

UNIVERSIDADE DE BRASÍLIA (UnB)  
FACULDADE DE COMUNICAÇÃO (FAC)  
DEPARTAMENTO DE JORNALISMO (JOR)

**JORNALISMO DE DADOS: CONCEITOS, FLUXOS,  
FERRAMENTAS E FORMAÇÃO ACADÊMICA**

LARISSA DE JESUS SILVA

BRASÍLIA - DF  
NOVEMBRO DE 2019

LARISSA DE JESUS SILVA

**JORNALISMO DE DADOS: CONCEITOS, FLUXOS,  
FERRAMENTAS E FORMAÇÃO ACADÊMICA**

Trabalho de Conclusão de Curso  
apresentado à Faculdade de  
Comunicação da Universidade de Brasília,  
como parte dos requisitos para obtenção  
do título de Bacharel em Jornalismo.

**Orientadora:** Márcia Marques

**Co-Orientador:** Benedito Medeiros Neto

BRASÍLIA - DF

2019

Dedico este trabalho aos meus pais e irmão, pelo apoio em todas as horas, pela compreensão e carinho infindável. Aos amigos e colegas que acompanharam minha trajetória e se alegraram com minhas vitórias, e a Deus pela proteção diária e provisão.

## **AGRADECIMENTOS**

Aos professores que participaram da banca de defesa deste trabalho de conclusão de curso, especialmente a Márcia Marques e Benedito Medeiros, que me orientaram ao longo do desenvolvimento do tema escolhido, desde o projeto experimental.

À minha colega de trabalho e amiga Andrea Matias Silva Mota pelos conselhos, compreensão e apoio constante, principalmente ao longo deste ano, e também pelo aceite do convite a participar da minha banca de defesa como suplente.

*“You can have data without information, but  
you cannot have information without data.”*

— Daniel Keys Moran

## RESUMO

A exploração e utilização de dados para fins jornalísticos não é uma prática nova, mas é evidente a expansão contínua e reconhecimento das práticas do Jornalismo de Dados em todo o mundo. Esse movimento, impulsionado pela popularização de termos como bases de dados e big data, conquistou espaço nas mesas de debates sobre o futuro do jornalismo e das transformações no perfil dos profissionais desta área. Este trabalho tem como objetivo entender porque é importante falar sobre Jornalismo de Dados (JD), apresentando detalhes sobre seu surgimento, estruturação e expansão. Também discorreremos sobre os conceitos e tipos de dado, bases de dados, leis de acesso à informação e suas relações com a apuração jornalística. Por fim, analisamos a presença do tema, enquanto disciplina, nas principais instituições acadêmicas do país que oferecem cursos de graduação em Jornalismo. A análise mostra a atual situação da inserção dos conhecimentos e técnicas do Jornalismo de Dados nos currículos obrigatórios do ensino superior brasileiro. Observamos que ainda é baixa a oferta de disciplinas sobre dados na graduação, algo que evidencia o ritmo lento no qual os avanços da profissão no país caminham, e também define um ponto de atenção para a falta de atualização das matrizes curriculares das faculdades, que as impedem de acompanhar o ritmo de transformações e inovações do mercado de trabalho. Entretanto, o resultado da análise ainda é positivo. Mesmo que em passos curtos, devemos reconhecer que temos avançado na tentativa de seguir as tendências do mercado, de formar profissionais cada vez mais qualificados para lidar com a nova realidade hiper tecnológica na qual todos estamos inseridos.

**Palavras-chave:** jornalismo de dados, ferramentas, fluxos de trabalho, formação acadêmica.

## **ABSTRACT**

Exploring and using data for journalistic purposes is not a new practice, but it is evident that there is continued expansion and recognition of data journalism practices around the world. This movement, driven by the popularization of terms such as databases and big data, gets space in the debate tables about the future of journalism and the changes in the profile of professionals in this area. This paper aims to understand why it is important to talk about Data Journalism (DJ), details about its emergence, structuring and expansion. We will also discuss the concepts and types of data, databases, laws on access to information and their relationship with journalistic investigation. Finally, we analyze the presence of the subject, as discipline, in the main academic institutions of the country that offer undergraduate courses in Journalism. The analysis shows the current situation of the insertion of Data Journalism knowledge and techniques in the obligatory curricula of Brazilian higher education. We note that there is still a low supply of data subjects at undergraduate level, something that shows the slow pace and there are no advances in the profession in the country. the accompaniment or rhythm of changes and innovations in the labor market. However, the result of the analysis is still positive. Even if the steps are small, it's important to recognize that we have been trying to follow market trends and training professionals to deal with this new hyper technological reality where we are all inserted.

**Keywords:** data journalism, tools, workflow, academic background.

## LISTA DE FIGURAS

Figura 1 - Visualização de dados da reportagem "How Brexit revealed four new political faction"	26
Figura 2 - Infográfico da reportagem "Cuadernos: La Investigación Que Cambió La Historia"	28
Figura 3 - Atual visualização do Basômetro	29
Figura 4 - Exemplo de infográfico desenvolvido para resultados da reportagem do Estadão Dados	30
Figura 5 - Exemplo de infográfico desenvolvido pela DeltaFolha para o projeto GPS Eleitoral	32
Figura 6 - Fluxo de trabalho de Lorenz para Jornalismo de Dados	48
Figura 7 - Pirâmide Invertida de Bradshaw	49
Figura 8 - Workflow de Rogers para JD	50
Figura 9 - Fluxo do Jornalismo de Dados por Silver	52
Figura 10 - Sugestão de Fluxo de Produção para Jornalismo de Dados	53
Figura 11 - Etapa de coleta de dados do workflow de Bradshaw	54
Figura 12 - Matriz de habilidades e competências no Jornalismo de Dados brasileiro	64
Figura 13 - Número de instituições nacionais que oferecem graduação em Jornalismo e	67



## **LISTA DE TABELAS**

Tabela 1 - Cinco Porquês para Jornalismo de Dados	51
Tabela 2 - Ferramentas do Google para Etapas do Workflow do JD	59
Tabela 3 - Mapeamento com quinze maiores instituições nacionais de ensino superior	65
Tabela 4 - Mapeamento das instituições de ensino superior com graduação em Jornalismo	68

## SUMÁRIO

INTRODUÇÃO .....	10
CAPÍTULO 1 - JORNALISMO DE DADOS .....	13
1.1 História e Conceitos .....	13
1.2 Organizações Apoiadoras do Jornalismo de Dados .....	21
1.3 Principais Núcleos de Jornalismo de Dados .....	24
CAPÍTULO 2 - DADOS, INFORMAÇÃO E CONHECIMENTO .....	36
2.1 Tipos de Dados, Big Data e Bases de Dados .....	38
2.2 Dados Abertos .....	45
CAPÍTULO 3 - FLUXO DE PRODUÇÃO E AS FERRAMENTAS .....	50
3.1 Sobre os Processos .....	50
3.2 Sobre as Ferramentas e Habilidades .....	57
3.2.1 Google e Ferramentas para Jornalismo de Dados .....	60
3.3 Sobre Representação dos Dados e Infografia .....	65
CAPÍTULO 4 – ENSINO DE JORNALISMO DE DADOS NO BRASIL .....	68
4.1 Faculdades de Jornalismo no Brasil .....	70
4.2 Faculdades de Jornalismo no Distrito Federal .....	72
4.3 Aparentamentos sobre os Resultados Observados .....	74
CONSIDERAÇÕES.....	76
REFERÊNCIAS .....	80

## INTRODUÇÃO

A exploração e utilização de dados para fins jornalísticos não é uma prática nova. Toda atividade de apuração jornalística é estruturada com base em dados, ou informações, previamente existentes, que possibilitem o desenvolvimento de uma narrativa e que possam contar uma história, inédita ou não. Entretanto, quando nossos objetos de apuração são grandes quantidades de dados digitais, ou seja, bases sistematizadas e organizadas, ou até mesmo compilados desorganizados de informações que demandam manipulação por meio de computadores e softwares especializados, o método de apuração assim como o produto final — a reportagem — representam uma nova forma, ou maneira de se fazer jornalismo que integra desde as técnicas mais tradicionais até competências específicas e multidisciplinares.

Essa necessidade em introduzir novas práticas às rotinas jornalísticas é resultado do surgimento e desenvolvimento da Internet, que também está relacionada à integração massiva de Tecnologias de Informação e Comunicação (TICs) em diversos setores da sociedade contemporânea (LIMA JUNIOR, 2012, p.208). As informações que antes só poderiam ser encontradas em formatos físicos, em arquivos, bibliotecas e acervos, passaram a estar disponíveis na rede, como dados digitais, sendo armazenados enquanto novos dados são constantemente gerados dentro das próprias redes e suas conexões, atingindo então volumes imensuráveis de partículas de informação.

Este cenário é um fator de grande influência no crescimento e fortalecimento do Jornalismo de Dados. Obtendo, selecionando e apresentando informações que não estão explícitas, sob difícil acesso e compreensão do cidadão leitor comum. Se a linguagem desta sociedade hiperconectada é o dado, o jornalista exerce então um papel de tradutor; aquele que obtém, trata, checa e expõe os assuntos de interesse público. Assim, em expansão contínua e sob demanda progressiva, as técnicas de apuração e produção de reportagens com dados estão ganhando espaço e se consolidando nas mesas de debate sobre jornalismo da era digital, assim como

estão cada vez mais presentes nas redações profissionais, desde as mais tradicionais até as mídias independentes e especializadas.

Por fim, este trabalho de conclusão de curso em formato de monografia de análise teórica, construída sob metodologia de pesquisa bibliográfica (LIMA e MIOTO, 2007, p. 38; GERHARDT e SILVEIRA, 2009, p. 37), apresentará uma introdução ao Jornalismo de Dados, por meio da discussão dos principais conceitos e autores que tratam sobre a história e desenvolvimento do assunto no mundo. O objetivo deste trabalho é ressaltar a importância do tema, entender seu surgimento e expansão, analisar a presença do JD nas principais instituições acadêmicas do país para então respondermos a pergunta-problema que originou esta monografia: por que falar de Jornalismo de Dados? Para isto, dividimos este trabalho em quatro capítulos.

O primeiro capítulo tratará do surgimento e consolidação dos conceitos sobre o Jornalismo de Dados no mundo (MEYER, 1979; PAUL, 1999; GARRISON, 1998; COX, 2000; HOLOVATY, 2006; BOUNEGRU, 2012; TRASEL, 2014 e outros), assim como também apresentará as organizações e instituições que têm apoiado e financiado o crescimento deste jornalismo, e os grandes centros de produção jornalística com dados, nacionais e internacionais.

A segunda seção, com caráter mais técnico, abordará conceitos paralelos importantes para o entendimento do trabalho com dados (KNUTH, 1996; ALBRECHT e OHIRA, 2000; ALMEIDA, 2002; CORREIA, 2009; SANTO, 2009; RAMAKRISHNAN e GEHRKE, 2011; KORTH, HENRY e SUDARSHAN, 2012; MARTINS, SIMÕES E SÁ, 2014; LIMA, 2017 e outros), enquanto que o terceiro capítulo falará estritamente sobre os fluxos de trabalho (*workflow*) do JD, abrangendo as ferramentas, técnicas e habilidades.

Por fim, a quarta e última parte trará apontamentos sobre a formação em Jornalismo no Brasil e a inserção do Jornalismo de Dados nos currículos de

graduação (MIELNICZUK e TRÄSEL, 2017; MOLINA e MEDEIROS, 2017; OLIVEIRA e ANGELUCI, 2019). Destacamos que, para a construção desta monografia, buscamos citar os autores principais e obras clássicas, mas também nos preocupamos em trazer referências mais recentes sobre o tema, que já analisassem e problematizassem aspectos recentes, como tecnologias e técnicas pertencentes ao cenário atual da sociedade.

## CAPÍTULO 1 - JORNALISMO DE DADOS

As propriedades tecnológicas da internet têm sido lapidadas por intermédio do aumento da velocidade de transmissão de dados, da interconexão entre máquinas computacionais com poderosa capacidade de processamento e armazenamento de dados, e da aplicação de sofisticadas linguagens de programação, que possibilitam a mineração e a análise de extensos e complexos bancos de dados — *big data*<sup>1</sup>.

Manuel Castells (2003) afirma que a rede de internet se tornou a estrutura essencial que suporta toda ordenação do que conhecemos hoje como a Era da Informação. Para o autor, a tecnologia da informação equivale ao que a eletricidade representou para a Era Industrial, no sentido de que “em nossa época a Internet poderia ser equiparada tanto a uma rede elétrica quanto ao motor elétrico, em razão de sua capacidade de distribuir a força da informação por todo o domínio da atividade humana” (CASTELLS, 2003, p. 7).

Influenciada por este cenário contemporâneo altamente digitalizado, a atuação do profissional de Jornalismo vem sendo significativamente transformada e refinada, devido à latente necessidade de aproximação com demais áreas do conhecimento, especificamente com o campo das Ciências e Tecnologias da Computação e da Informação. A conexão entre essas áreas é bastante favorável visto que o entendimento das lógicas computacionais viabiliza o uso e a descoberta de novos métodos para o tratamento de dados (informação), por meio das tecnologias, no sentido de intensificar o desenvolvimento de abordagens jornalísticas inovadoras. Essa fusão de inteligências, que já está acontecendo, é que tem facilitado o desenvolvimento do chamado Jornalismo de Dados (JD) ou Jornalismo Guiado por Dados (JGD), que neste trabalho serão referidos como sinônimos; estes representam uma vertente importante que está cada vez mais popular e presente nas redações.

---

<sup>1</sup>*Big Data* é um conjunto de dados maior e mais complexo, especialmente de novas fontes de dados. Esses conjuntos de dados são tão volumosos que softwares tradicionais de processamento de dados simplesmente não conseguem gerenciá-los. São dados com maior variedade que chegam em volumes crescentes e com velocidade cada vez maior

## 1.1 História e Conceitos

Ainda que o Jornalismo de Dados tenha ganhado força recentemente, Mancini e Vasconcellos (2016) defendem que o termo provém de práticas que tiveram início na década de 1960, pelo jornalista Philip Meyer, com o conceito de Jornalismo de Precisão. Em 1967, trabalhando temporariamente para a revista *Detroit Free Press*, Meyer produziu uma reportagem sobre os conflitos raciais que borbulhavam na cidade de Detroit na época, utilizando um computador *IBM 360 mainframe* para analisar dados demográficos, provenientes de um questionário aplicado a uma amostra representativa dos moradores dos bairros afetados, para investigar a identidade e o comportamento daquelas pessoas (VENTURA, 2018, p. 243).

O objetivo era entender se a classe econômica e o nível de escolaridade influenciavam ou não as pessoas em participar dos motins. A reportagem, que foi intitulada *The People Beyond 12th Street*, vencedora do Prêmio Pulitzer<sup>2</sup> no mesmo ano de sua publicação, mostrou que os fatores qualitativos analisados não estavam atrelados à escolha de participar ou não dos conflitos (TRÄSEL, 2014, p. 98). Assim, o Jornalismo de Precisão de Meyer tratava da inserção do computador, associado com as metodologias das ciências sociais, na produção de reportagens com altos índices de exatidão (VENTURA, 2018).

Com a aplicação prática desse conceito, Philip Meyer desenvolveu uma forma inovadora de otimizar a produção da notícia, e de minimizar as chances de erro unindo as ciências sociais a utilização do computador na manipulação de informações, para que então fosse possível a análise dos dados obtidos e, assim, a estruturação de uma reportagem. Meyer foi o precursor do que anos depois Nora Paul chamou de *Computer-Assisted Reporting* (CAR), ou Reportagem Assistida por Computador (RAC) — também denominado *Computer-Assisted Journalism*, uma

---

<sup>2</sup>O Prêmio Pulitzer é oferecido a pessoas que realizam trabalhos de excelência nas áreas de jornalismo, literatura e música. A premiação foi criada em 1917, por Joseph Pulitzer, e hoje é administrada pela Universidade de Colúmbia em Nova Iorque. Os indicados ao prêmio são escolhidos por uma banca independente e concorrem em 21 categorias.

técnica de reportagem que utiliza computadores e softwares como ferramentas para a obtenção e análise de informações (PAUL, 1999 e GARRISON, 1998 apud BARBOSA, 2007a). Tais termos ainda são utilizados atualmente em fóruns especializados, como o *Investigative Reporters e Editors* (IRE) e a Associação Brasileira de Jornalismo Investigativo (Abraji).

Melisma Cox (2000) argumenta que a era da reportagem com auxílio de computador teve início em 1952, quando a rede de TV americana *CBS News* de Washington usou um computador para analisar a disputa presidencial dos Estados Unidos entre Dwight D. Eisenhower e Adlai Stevenson. As previsões apontavam para uma eleição bastante acirrada. O computador, um *Remington Rand UNIVAC*, foi utilizado para prever o resultado da eleição com base nos retornos iniciais. Programadores haviam preparado fórmulas baseadas nesses retornos para calcular uma previsão mais exata possível do que poderia ser o resultado final.

Apesar do pensamento generalizado de que a eleição seria pareada, as previsões antecipadas pelo cálculo do UNIVAC disseram que Eisenhower venceria com uma esmagadora maioria absoluta dos votos. A *CBS* estava relutante em divulgar as previsões aparentemente impossíveis, então interromperam as apurações e, quando enfim transmitiram as previsões, eles foram criticados por terem se recusado a acreditar no computador. Segundo Cox (2000, p. 6), a era da reportagem assistida por computador nasceu durante esta disputa presidencial. E é fato que nenhuma eleição desde então deixou de utilizar computadores como ferramenta de previsão dos resultados.

Entretanto, esse episódio com a *CBS* não representou o ponto inicial do Jornalismo de Dados. Esse entendimento é defendido por Marcelo Träsel, (2014, p. 28) quando afirma que o caso descrito por Cox foi algo isolado na época, e que não foi seguido de outras aplicações semelhantes envolvendo computadores e a produção de reportagens, sequer na mesma emissora e nem em outras redações. Em sua tese, o autor diz que definir um período exato para a implementação dessas



práticas no meio jornalístico em geral é difícil, visto que “diferentes redações seguiram diferentes ritmos de informatização, dependentes não apenas de sua cultura específica, mas também de fatores econômicos e do preparo da mão de obra” (TRÄSEL, 2014, p. 29).

Bruce Garrison (1998) defende que a popularização dos computadores pessoais nos anos 80 fez com que, gradativamente, as técnicas de Reportagem Assistida por Computador se tornassem mais presentes nas redações americanas. No final da década de 90, quando os computadores e as bases de dados (BDs) alcançaram um alto nível de utilização nas empresas jornalísticas, a partir da consolidação da RAC, e quando a *World Wide Web* havia se tornado a base de dados online e fonte de informação mais popular nas redações, Garrison deixou registrado assertivamente em seu livro *Computer-Assisted Reporting* que a era da base de dados havia chegado.

E assim, a reportagem com auxílio de computador se desenvolveu a tal ponto que as bases de dados se tornaram fundamentais para o trabalho dos jornalistas, pois fariam todo o diferencial nas reportagens e notícias, acrescentando mais qualidade de interpretação e de contextualização para os acontecimentos (BARBOSA, 2006, p. 4). Em 2001, os computadores atingiram uma massa crítica nas salas de redação americanas em termos de uso geral para pesquisas on-line, pesquisa de conteúdo não especializado e frequência diária de uso on-line, mostrando que a Reportagem Assistida por Computador havia se tornado onipresente nos Estados Unidos (GARRISON, 2001).

No Brasil, os *personal computers* (PCs), ou computadores pessoais, começaram a ser introduzidos nas redações a partir dos anos 80. O jornal *Folha de S. Paulo* foi o primeiro do país a introduzir terminais de computador — sucatas compradas de jornais americanos — em suas redações, projeto que vinha sendo elaborado desde 1968 e foi efetivado em 1983 (VIANNA, 1992 apud LIMA JUNIOR, 2008, p. 208). Todavia, delongou cerca de duas décadas para entrelaçar essas

tecnologias no cotidiano profissional do jornalista brasileiro. Anteriormente, no final dos anos 60 e início dos 70 do século passado, o *The New York Times* já havia estruturado o primeiro banco de dados, que foi inserido nas etapas de produção da notícia. A agência de notícias *Reuters*, em 1968, foi pioneira ao utilizar computadores nas conexões da sua rede interna para gerenciar a demanda de notícias recebidas (LIMA JUNIOR, 2008, p. 209).

É possível observar que, desde o contato preliminar com os computadores e a possibilidade de digitalização das rotinas jornalísticas, o Brasil esteve alguns passos atrás, e esse atraso também pode representar um retardamento no desenvolvimento do Jornalismo de Dados nas redações brasileiras. Discorreremos sobre este ponto de vista adiante.

Além destes elementos físicos — referindo à modernização de equipamentos e ferramentas das redações —, o próprio perfil do profissional de jornalismo também é uma importante influência-consequência das transformações que, mais tarde, arquitetaram as bases do Jornalismo de Dados. Philip Meyer afirmou na segunda edição de sua obra, *Precision Journalism* (1979) que os jornalistas interessados em RAC compunham uma linhagem particular de repórteres, que frequentemente compravam seus próprios computadores antes mesmo de as redações os adotarem.

Houve um tempo em que tudo o que se precisava era dedicação à verdade, muita energia e algum talento para escrever. Você ainda precisa dessas coisas, mas elas não são mais suficientes. O mundo ficou tão complicado, o aumento da informação disponível tão explosivo, que o jornalista precisa ser filtro e transmissor, organizador e intérprete, além de coletar e entregar fatos. Além de saber como colocar a informação na página ou no ar, também deve saber colocá-la na cabeça do receptor. Em resumo, um jornalista deve ser administrador de bases de dados, processador de dados e analista de dados (MEYER, 1979 apud O ESTADO DE S. PAULO, 2011, online<sup>3</sup>).

Todo este *background* do qual tratamos até o momento — abrangendo o desenvolvimento da internet, a digitalização das redações e os conceitos de

---

<sup>3</sup>Tradução feita pelo jornal *O Estado de S. Paulo* para a matéria *O novo novo jornalismo* publicada em 2011. Veja a íntegra: <<https://brasil.estadao.com.br/blogs/em-foca/o-novo-novo-jornalismo/>>

Jornalismo de Precisão e de Reportagem Assistida por Computador — cooperou para a estruturação de uma base para as concepções teóricas e práticas do objeto de estudo deste trabalho. Então, como surgiram os termos Jornalismo de Dados e Jornalismo Guiado por Dados?

Segundo Liliana Bounegru, para a primeira edição do *The Data Journalism Handbook* (2012, p. 18), a referência mais antiga a *Data Journalism* (Jornalismo de Dados), termo que derivou em *Data-Driven Journalism* (Jornalismo Guiado por Dados), foi feita pelo programador Adrian Holovaty no artigo *A fundamental way newspaper sites need to change*<sup>4</sup>, publicado em seu website pessoal em 2006. Na publicação, Holovaty defende a inserção de técnicas de gerenciamento de bases de dados ao cotidiano das redações, como forma de facilitar o reaproveitamento das informações coletadas no trabalho diário de reportagem — uma das primeiras formulações daquilo que está mais próximo do que é hoje considerado o Jornalismo de Dados.

Por exemplo, digamos que um jornal escreveu uma história sobre um incêndio local. A possibilidade de ler essa notícia em um celular é bem legal. Obrigada, tecnologia! Mas o que eu realmente quero fazer é explorar os fatos, um por um, com camadas de atribuição e uma infraestrutura para comparar os detalhes do incêndio com os detalhes de incêndios anteriores: data, hora, local, vítimas, número do corpo de bombeiros, distância do corpo de bombeiros, nomes e anos de experiência dos bombeiros no socorro, tempo necessário para a chegada dos bombeiros e incêndios subsequentes, sempre que eles aconteçam (HOLOVATY, 2006, online, tradução da autora).

Desde Holovaty e suas primeiras concepções do que poderia ser este jornalismo preocupado com dados, muitos outros estudiosos da área surgiram com diversos conceitos e nomenclaturas no intuito de definir e delimitar o que é e como fazer esse tipo de jornalismo. Algumas definições se concentram no propósito do Jornalismo de Dados, como sendo a combinação do tradicional ‘faro por notícias’ com a escala e o alcance da informação digital agora disponível. Outros são mais direcionados aos processos que ajudam a produzi-lo: “coleta, limpeza, organização,

---

<sup>4</sup>Um aspecto fundamental em que os websites de jornais precisam mudar. Íntegra da publicação: <<http://www.holovaty.com/writing/fundamental-change/>>.

análise, visualização e publicação de dados para apoiar a criação de atos de jornalismo” (HOWARD, 2014 apud GOOGLE NEWS LAB, 2017, online, tradução da autora).

Sandra Crucianelli argumenta que o Jornalismo de Dados está extensivamente ligado às práticas jornalísticas como o jornalismo de investigação, o jornalismo de precisão e o jornalismo de profundidade, ou Reportagem Assistida por Computador (RAC), confundindo-se, por vezes com algumas delas. Para Crucianelli (2013 apud MARTINHO, 2014, p. 69), aquilo que na atualidade se nomina de Jornalismo de Dados, envolve todas as práticas citadas, podendo envolver mais ou menos cada uma delas, dependendo dos tipos de matérias que serão produzidas. O aparecimento do JD só é possível com o amadurecimento de algumas destas práticas jornalísticas.

Primeiramente, para que um trabalho jornalístico possa ser considerado de investigação são necessárias três condições (CRUCIANELLI, 2013 apud MARTINHO, 2014, p. 69): que trabalhe um tema de relevância social, que alguém tenha interesse em manter segredo sobre a matéria, e que pressuponha da parte do jornalista um trabalho adicional. Um trabalho de jornalismo de dados também pode, por isso, ser jornalismo de investigação. No caso do jornalismo de profundidade, o trabalho tem os mesmos princípios que o jornalismo de investigação, exceto a premissa do segredo, que aqui não se aplica. O jornalismo de precisão baseia-se no argumento de que os jornalistas devem utilizar os métodos de investigação usados nas ciências sociais, recorrendo à investigação qualitativa e à investigação quantitativa.

E, por vezes, quando se faz jornalismo de dados também são aplicados alguns destes métodos. Crucianelli define a Reportagem Assistida por Computador como a utilização de recursos computacionais para a recolha e análise de dados, chegando mesmo a afirmar que quando se fala do tratamento de grandes

quantidades de dados, já não se deve falar de RAC, mas sim de Jornalismo de Dados.

Jonathan Gray, da *Open Knowledge Foundation*, argumenta que Jornalismo de Dados é um termo bastante recente usado para descrever um conjunto de práticas que usam dados para melhorar as notícias. No artigo *The future of Data Journalism* publicado em 2012, Gray explica que essas práticas incluem o uso de bancos de dados e ferramentas analíticas para escrever melhores histórias e fazer melhores investigações, a publicação de conjuntos de dados relevantes ao lado de histórias, e o uso de conjuntos de dados para fornecer visualizações interativas de dados ou aplicativos de notícias.

Bounegru (2012), para o *The Data Journalism Handbook 1*, defende a importância do uso de tais ferramentas, pois inserem o JD no ecossistema de artefatos e práticas que brotam em sites de dados e serviços, pois, segundo ela, as tecnologias digitais e a web estão fundamentalmente modificando a forma com que a informação é publicada. Bounegru também afirma que o Jornalismo de Dados é o conhecimento de dados em massa. Simpatizante deste mesmo ponto de vista, o inventor do WWW (*World Wide Web*), Tim Berners-Lee, em entrevista para o *The Guardian*, observou que o trabalho com dados é o futuro para os jornalistas. Para ele, os jornalistas necessitam ser especialistas em dados.

No Brasil, embora com algumas nuances distintas, os autores, entre eles pesquisadores e jornalistas que trabalham com Jornalismo de Dados, adotam uma visão que associa a capacidade investigativa e as novas possibilidades da tecnologia para conceituar esse campo jornalístico. Segundo Marcelo Träsel (2017, p. 2), essa forma de se fazer jornalismo tem como objetivo “a produção, tratamento e cruzamento de grandes quantidades de dados, de modo a permitir maior eficiência na recuperação de informações e na apuração de reportagens a partir de conjuntos de dados”.

Para Träsel, as técnicas do JGD viabilizam a localização de possíveis pautas de interesse, no meio de extensos conjuntos de dados com inúmeras informações, que seriam impossíveis de manipular sem o suporte de computadores. Assim como também possibilitam a execução de comparações entre bases de dados distintas com o intuito de perceber padrões e discrepâncias, produzindo novos conhecimentos.

Como apresentado ao longo deste capítulo, até este ponto, não há uma definição clara e universal do que seria o Jornalismo de Dados. Os conceitos existentes são combinações de outras práticas e técnicas jornalísticas paralelas, que se cruzam na execução de reportagens baseadas (guiadas) em dados. Contudo, ainda na tentativa de consolidar uma definição para o JD, um estudo realizado em 2017 pelo *Google News Lab* em parceria com o *PolicyViz*, trouxe uma classificação que ajuda a estruturar um entendimento simples do Jornalismo de Dados com base nos tipos de notícias produzidas a partir dele, sendo estas: notícias que são enriquecidas por dados, notícias que usam dados para investigar histórias e notícias que explicam dados.

Apesar da formulação dessas classificações, com base em entrevistas qualitativas — conduzidas com cerca de novecentos jornalistas ao redor do mundo — uma das principais conclusões deste estudo foi que “existe uma implícita, se não explícita, expectativa de que as organizações de notícias usarão dados nas suas reportagens” (GOOGLE NEWS LAB, 2017, p. 22, tradução da autora).

## **1.2 Organizações Apoiadoras do Jornalismo de Dados**

É de extrema importância que reconheçamos o papel e influência dos profissionais, veículos de comunicação e instituições diversas que cooperaram para a expansão do Jornalismo de Dados, por meio da defesa deste campo de estudo, pela atuação no desenvolvimento de reportagens com dados e pelo investimento direto em projetos e programas que fizeram com que este tipo de jornalismo

consolidasse seu espaço no meio profissional, no mercado midiático e também no meio acadêmico.

Entretanto, os atores e organizações são numerosos e se precisássemos citar todos em suas individualidades esta monografia triplicaria sua quantidade de páginas. Com isto em mente, trataremos a seguir sobre os principais responsáveis — ou os mais recorrentes e citados na literatura especializada — mundiais e nacionais pela estruturação e fortalecimento do Jornalismo de Dados nas últimas duas décadas.

Um dos projetos mais importantes e influentes na história do JD no mundo talvez seja a fundação da *WikiLeaks*, uma plataforma online criada em 2006 com o objetivo de enviar documentos de todo tipo, de forma anônima e teoricamente segura para análise e publicação. Entretanto, o projeto só se tornou jornalisticamente relevante quando as informações brutas ali dispostas passaram a ser lapidadas, filtradas e organizadas por jornalistas de órgãos de comunicação social de alguns dos maiores grupos midiáticos mundiais.

Em 2010, com o vazamento dos chamados *Iraq War Logs*<sup>5</sup>, a *WikiLeaks* se firmou como um incentivador de muitos questionamentos, que foram fundamentais para o Jornalismo de Dados, relacionados à competência dos veículos de comunicação para analisar documentos daqueles tipos e volumes, sobre a confiabilidade na origem daqueles dados, sobre a legalidade na divulgação de informações de guerra e muitas outras interrogações semelhantes (MARTINHO, 2014, p. 66).

Alguns anos depois, em 2012, duas grandes e importantes organizações ainda atuantes no desenvolvimento do jornalismo no mundo, o *European Journalism Centre* (EJC) e a *Open Knowledge Foundation* (OKF) concluíram o primeiro guia que

---

<sup>5</sup>Os documentos divulgados pela *WikiLeaks* sobre a Guerra do Iraque ainda estão disponíveis para acesso e continuam sendo atualizados. Veja mais sobre: <<https://wardiaries.wikileaks.org>>.

sistematizou um entendimento básico e primordial sobre o Jornalismo de Dados. O *The Data Journalism Handbook*, que já citamos anteriormente, é um documento construído colaborativamente, que reúne exemplos e análises de especialistas de vários países sobre o tema.

Rapidamente, o manual foi traduzido do inglês para o espanhol e então para o chinês e, em 2014, foi lançado em português, graças ao esforço da Associação Brasileira de Jornalismo Investigativo (Abraji), que é uma das instituições mais atuantes na defesa e propagação do Jornalismo de Dados no Brasil. Atualmente já está disponível para leitura on-line a segunda edição do *The Data Journalism Handbook*, que conta agora também com o apoio e patrocínio da *Google News Initiative*.

A *Google News Initiative* (GNI), foi lançada pela Google com o objetivo de impulsionar o avanço tecnológico nas empresas jornalísticas. Presente no Brasil e em outros 70 países, a iniciativa já apoiou 350 organizações com um investimento de US\$ 120 milhões em ações e projetos, segundo Marco Túlio Pires, coordenador do *Google News Lab* Brasil (THINK WITH GOOGLE, 2019, online). Antes mesmo da consolidação da GNI no Brasil, a Google já havia investido, somente em 2017, R\$ 10 milhões em veículos nacionais de comunicação e treinado 10 mil jornalistas.

Já o *Google News Lab* é uma equipe da GNI que trabalha mais diretamente com inovação na produção de notícias, oferecendo parcerias e treinamentos em cerca de 50 países. Como visto, a *Google News Initiative* atua fortemente ao redor do mundo na propagação dos conceitos e práticas de tecnologia e mundo digital para o jornalismo, principalmente no desenvolvimento das áreas de fact-checking, no combate às fake news, e no Jornalismo de Dados — dissertaremos sobre a GNI e o desenvolvimento de ferramentas para JD no próximo capítulo.

Citamos também o Centro Knight para o Jornalismo nas Américas da Universidade do Texas em Austin, que é um programa de extensão e capacitação



profissional para jornalistas na América Latina e no Caribe, que vem dedicando muitos de seus cursos ao estudo das técnicas do Jornalismo de Dados. Em atuação desde 2003, o *Knight Center* colaborou inclusive na criação da Abraji e de outros centros e iniciativas ao redor do mundo.

Outras iniciativas que podemos referenciar nos estudos e propagação mundial do JD, com cursos, congressos, publicações e parcerias são: o *International Center for Journalists* (ICJ), o *Nieman Lab*, o *Global Investigative Journalism Network* e o *Global Editors Network*, que organiza anualmente o *Data Journalism Awards* desde 2012 com o apoio do *Google News Lab*, *Knight Foundation* e patrocínio da *Microsoft* e *Chartbeat*. Todas as organizações citadas disponibilizam conteúdos (cursos, artigos, livros etc) traduzidos para o português brasileiro.

### 1.3 Principais Núcleos de Jornalismo de Dados

Neste tópico apresentaremos os principais veículos de imprensa que vêm produzindo conteúdo jornalístico com dados no Brasil e no mundo nos últimos anos. No início deste capítulo citamos alguns exemplos clássicos de jornais tradicionais americanos — Meyer para o *Detroit Free Press*, *CBS News* nas eleições de 1952 e *The New York Times* — pioneiros na utilização de dados na produção de suas reportagens, o que na época ainda era compreendido como Reportagem Assistida por Computador (RAC). Mas há três grandes veículos que se consolidaram com plataformas digitais na Web e que são considerados vanguardistas no movimento do Jornalismo de Dados internacional: *BBC* (UK), *The Guardian* (UK) e o *La Nación* (ARG).

A *BBC*, ou *British Broadcasting Corporation*, uma tradicional rede pública de imprensa do governo do Reino Unido, que há duas décadas têm inserido narrativas baseadas em dados, assim como infográficos e outros recursos, em suas reportagens. A produção mais antiga utilizando dados feita pelo veículo foi publicada em 1999, e tratava de uma tabela de classificação escolar que usava dados publicados pelo governo, na qual os leitores podiam encontrar escolas locais

inserindo um código postal e compará-las em vários indicadores. Atualmente a corporação possui uma equipe chamada *BBC Visual and Data Journalism* que produz todos os conteúdos de dados publicados no site de notícias da *BBC* e que também alimenta uma página na plataforma *Medium* onde, desde o começo deste ano, discorrem sobre aspectos mais técnicos do processo produtivo adotado pela redação.

Em 2009, o *The Guardian* lançava o projeto *Datablog*, uma página no website do jornal dedicada à publicação e análise de dados. Simon Rogers, fundador do projeto e um dos nomes mais influentes na estruturação de conceitos e técnicas do JD, descreveu recentemente, para uma matéria memorial publicada este ano pelos atuais editores de dados do jornal — Caelainn Barr, Mona Chalabi e Nick Evershed —, como foi o processo de desenvolvimento do projeto.

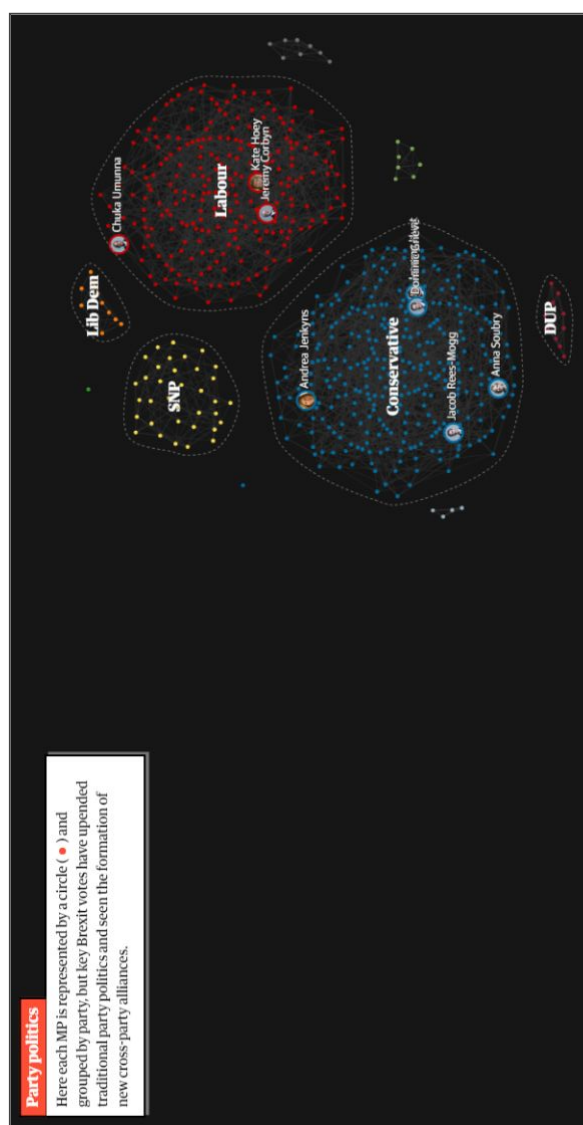
Segundo Rogers, a ideia fundamental que o impulsionou a iniciar o *Datablog* foi a de publicar os dados em posse do jornal em um formato que outras pessoas pudessem reutilizar os conteúdos. “Eu estava trabalhando como editor de notícias na equipe de gráficos e tinha muitos conjuntos de dados. Isso ocorreu no momento em que subitamente havia uma carga de dados sendo publicada, mas geralmente em formatos terríveis, como PDFs”, conta Rogers (BARR, CHALABI e EVERSLED, 2019, online, tradução da autora).

Desde 2009, o *The Guardian* foi responsável pela publicação de centenas de importantes reportagens de dados com diversas temáticas. Na edição deste ano do *Data Journalism Awards*, o jornal britânico foi finalista em duas categorias: *Best Data Journalism Team Portfolio — Large Newsroom* (Melhor Portfólio de Equipe de Jornalismo de Dados — Grande Redação) e *Data Visualisation of The Year* (Visualização de Dados do Ano) pela reportagem *How Brexit revealed four new political factions* (Como o Brexit revelou quatro novas facções políticas).

[...] Fomos o primeiro blog sobre dados em qualquer lugar da mídia convencional. Antes do lançamento do *FiveThirtyEight* em 2014, antes do

resultado, havia o Datablog. Através de uma combinação de grandes histórias - os registros de guerra do *WikiLeaks*, a cobertura de distúrbios de 2011 e o crowdsourcing de despesas dos parlamentares - realmente decolou. De repente, havia dados em todos os lugares, e os explicamos e os disponibilizamos. Numa época em que a confiança no jornalismo estava diminuindo, estávamos dizendo: "Aqui estão os dados, aqui estão as ferramentas que usamos e aqui está o resultado. Agora você tem uma chance." Eu realmente acredito que, hoje, qualquer um pode fazê-lo (BARR, CHALABI e EVERSHED, 2019, online, tradução da autora).<sup>6</sup>

Figura 1 - Visualização de dados da reportagem "How Brexit revealed four new political faction"



Fonte: *The Guardian* (2019)

<sup>6</sup> [...] We were the first blog about data anywhere in the mainstream media. Before the 2014 launch of FiveThirtyEight, before the Upshot, there was the Datablog. Through a combination of big stories – the WikiLeaks war records, the 2011 riots coverage and the MPs’ expenses crowdsourcing – it really took off. Suddenly, there was data everywhere, and we explained it and made it more available. At a time when trust in journalism was diminishing, we were there saying: “Here’s the data, here are the tools we used and here’s the result. Now you have a go.” I really believe that, today, anyone can do it.

Com base em Buenos Aires (Argentina), o jornal *La Nación*, fundado em 1870, se lançou no universo de produção jornalística com dados em 2011 com o *La Nación Data*, uma iniciativa com foco em dados abertos e divulgação de informações públicas sempre atualizadas. Segundo editorial publicado em março deste ano, escrito pela jornalista Bianca Pallaro, a missão estratégica deste núcleo de dados é gerar impacto que resulte em mudanças na Argentina por meio do uso e promoção do jornalismo com dados abertos (PALLARO, 2019, online, tradução da autora), ou seja, além do serviço tradicional de um jornal — que é trazer informação de qualidade —, o *La Nación* também procura conscientizar seus leitores, a população argentina, acerca do seu direito à informação garantido pelo Estado democrático.

Todos os dias, atualizamos nossos indicadores de séries de dados que estão vinculados às visualizações incorporadas nos artigos do dia. Todos esses infográficos interativos têm um botão de download para que o usuário reutilize as informações ou simplesmente conceda acesso ao banco de dados histórico em um formato aberto. Temos aproximadamente mais de 200 bancos de dados que frequentemente atualizamos e abrimos ao público [...] Somos usuários pesados de dados abertos e acesso a informações públicas (FOIA) na Argentina, a fim de ajudar na participação dos cidadãos e abrir o governo, porque acreditamos que isso melhora a qualidade da democracia. O primeiro passo para tornar famosos os conjuntos de dados é tornar-se USUÁRIO de dados abertos. E se não houver dados abertos, nos tornamos PRODUTORES e transformamos formatos fechados em dados abertos (PALLARO, 2019, online, tradução da autora).<sup>7</sup>

O *La Nación*, que já acumula diversos prêmios, ganhou também premiação pela Google na categoria Melhor Portfólio de Equipe de Jornalismo de Dados (*Best Data Journalism Team Portfolio - Large Newsroom*) no último *Data Journalism Awards*. Considerados *watchdogs*<sup>8</sup> em seu país, a equipe de dados do jornal leva em seu currículo a produção de grandes reportagens com dados sobre corrupção

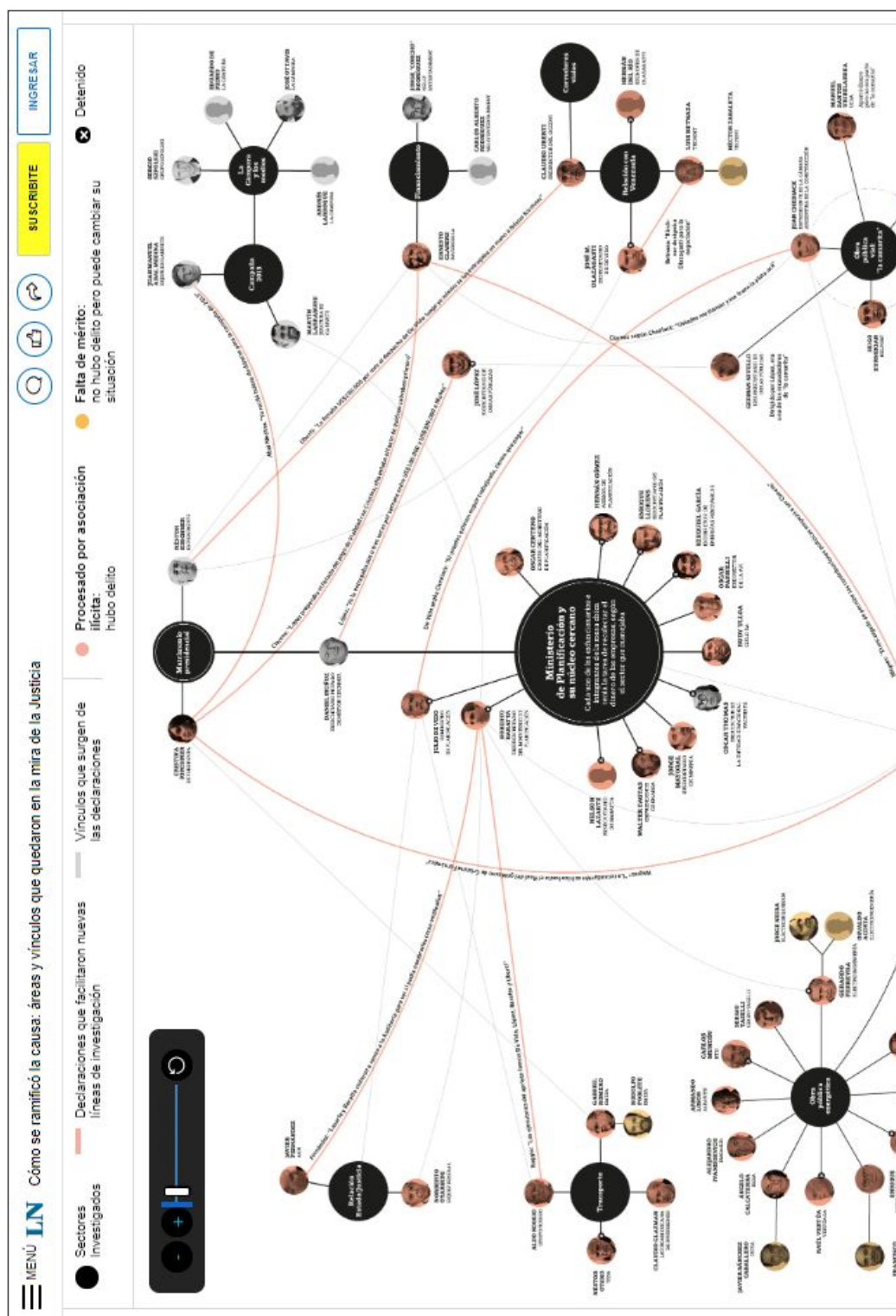
---

<sup>7</sup> Every day we update our data series indicators that are linked to visualizations that we embed in the articles of the day. All of this interactive infographics have a download button for the user to reuse the information or simply to give access to the historical database in an open format. and source. We have approximately more than 200 databases that we frequently update and open to the public. We are heavy users of Open Data and access to public information (FOIA) in Argentina, in order to help citizen participation and open government because we believe this improves the quality of democracy. The first step to make datasets famous is to become open data USERS. And if there is no open data, we become PRODUCERS and transform closed formats to open data.

<sup>8</sup>*Watchdogs*, ou cães de guarda, é uma expressão utilizada para descrever jornalistas investigativos que atuam na cobertura dos governos. O termo está relacionado ao conceito de 'Quarto Estado' atribuído à imprensa onde o jornalista figuraria como um guardião (ALBUQUERQUE, 2010, p. 92-93).

que ajudaram a identificar e punir legalmente atores que atentavam contra a transparência governamental argentina.

Figura 2 - Infográfico da reportagem "Cuadernos: La Investigación Que Cambió La Historia"



Fonte: La Nación (2019)

Outros veículos internacionais grandes também estão se destacando recentemente pela boa execução de reportagens com dados, os quais inclusive já foram premiados por suas produções em Jornalismo de Dados, como: *The Wall Street Journal* (EUA), *Reuters* (EUA), *South China Morning Post* (Hong Kong), *Caixin* (China) e *Al Jazeera* (Qatar). Assim como veículos menores e fundados recentemente: *Bloomberg* (EUA), *The Pudding* (EUA), *Postdata.club* (Cuba) e o *Media Hack Collective* (África do Sul).

No Brasil, o *Estado de S. Paulo* e a *Folha de S. Paulo* foram os pioneiros na integração em suas redações, inicialmente, das técnicas de RAC, que alicerçaram a futura introdução do Jornalismo de Dados nas rotinas desses jornais. No caso do jornal *O Estado de S. Paulo* — também conhecido como *Estadão*, fundado em 1875 — o núcleo de dados foi criado graças a uma parceria com o Centro Knight para o Jornalismo nas Américas e ao interesse do jornalista José Roberto de Toledo pela nova prática.

Assim, em 2012, a editoria *Estadão Dados* foi lançada com uma equipe de quatro pessoas, incluindo Toledo (TRÄSEL, 2014, p. 133). A descrição contida no site, na página de apresentação da editoria, diz que este é um “núcleo especializado em reportagens baseadas em estatísticas e no desenvolvimento de projetos especiais de visualização de dados” (SOBRE, 2019, online).

Aqui você vai encontrar gráficos e animações sobre temas do noticiário do dia, além de cruzamentos de dados e análises especiais feitas pela nossa equipe, formada por jornalistas e programadores. Os posts são divididos em três seções: gráfico do dia, permanentes, para séries estatísticas que serão constantemente atualizadas, e séries especiais, focadas em um tema específico (SOBRE, 2019, online).

Um dos projetos mais antigos e exitosos do *Estadão Dados* é o Basômetro, uma plataforma on-line que apresenta o desempenho de atores políticos em votações de projetos de lei, com atualizações feitas constantemente pela equipe.

A partir da percepção da disponibilidade dos dados de votações do Congresso em formato aberto e de um conhecimento prático sobre os

meandros da política brasileira, o ED [*Estadão Dados*] usou a inspiração do projeto desenvolvido pelo PoliGNU para conceber o Basômetro. Trata-se de um aplicativo baseado na Web que permite visualizar o comportamento de deputados federais e senadores da República a cada votação de projetos de lei desde o ano de 2003. [...] O banco de dados é atualizado manual e periodicamente a partir dos resultados de votações divulgados nos websites da Câmara dos Deputados e do Senado (TRÄSEL, 2014, p. 136).

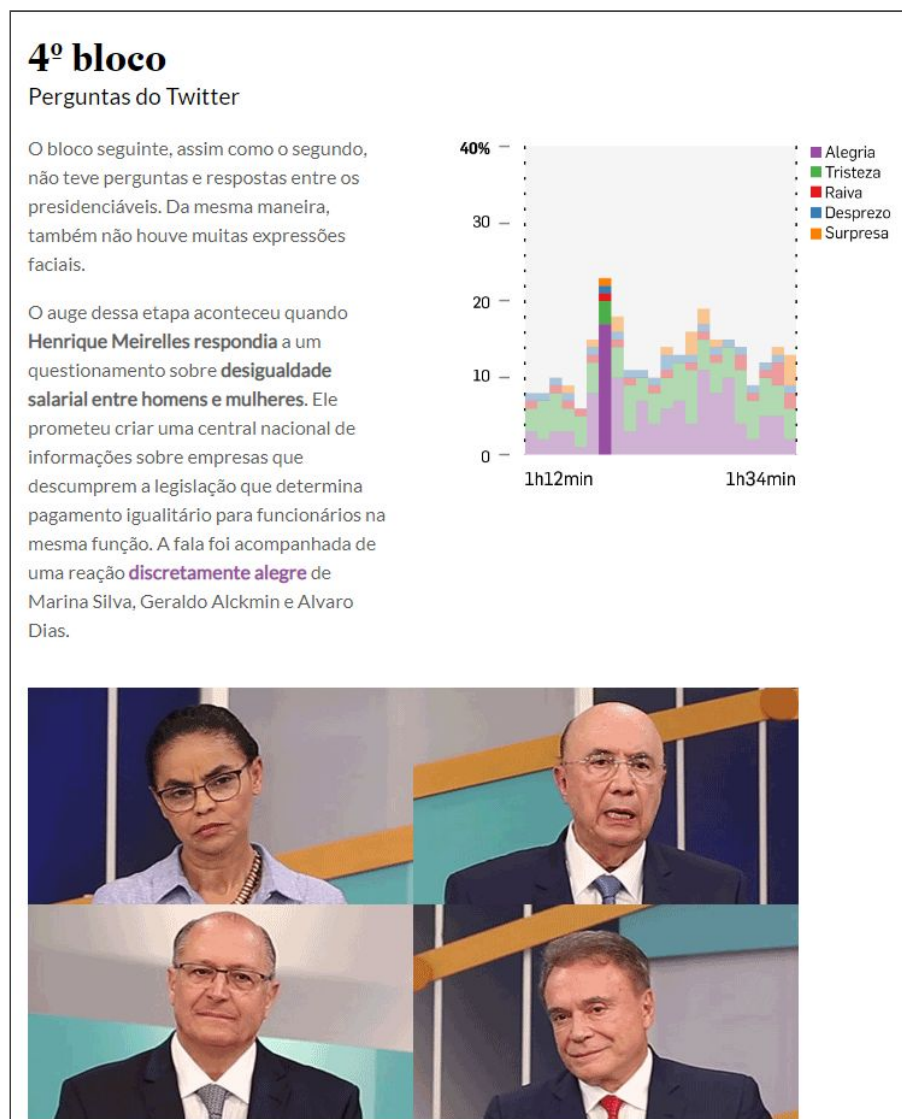
O Estadão Dados foi um dos indicados no *Data Journalism Awards* 2019, na categoria *Innovation in Data Journalism* (Inovação em Jornalismo de Dados), com a reportagem *O que revela uma análise das emoções dos candidatos durante o debate*, do repórter de dados Rodrigo Menegat, que trata de uma cobertura das eleições presidenciais brasileiras de 2018 onde, por meio da leitura facial dos rostos dos candidatos utilizando uma API, foi possível capturar e estudar as emoções dos presidencialistas ao longo dos debates.

Figura 3 - Atual visualização do Basômetro



Fonte: O Estado de S. Paulo (2019)

Figura 4 - Exemplo de infográfico desenvolvido para resultados da reportagem do Estadão Dados



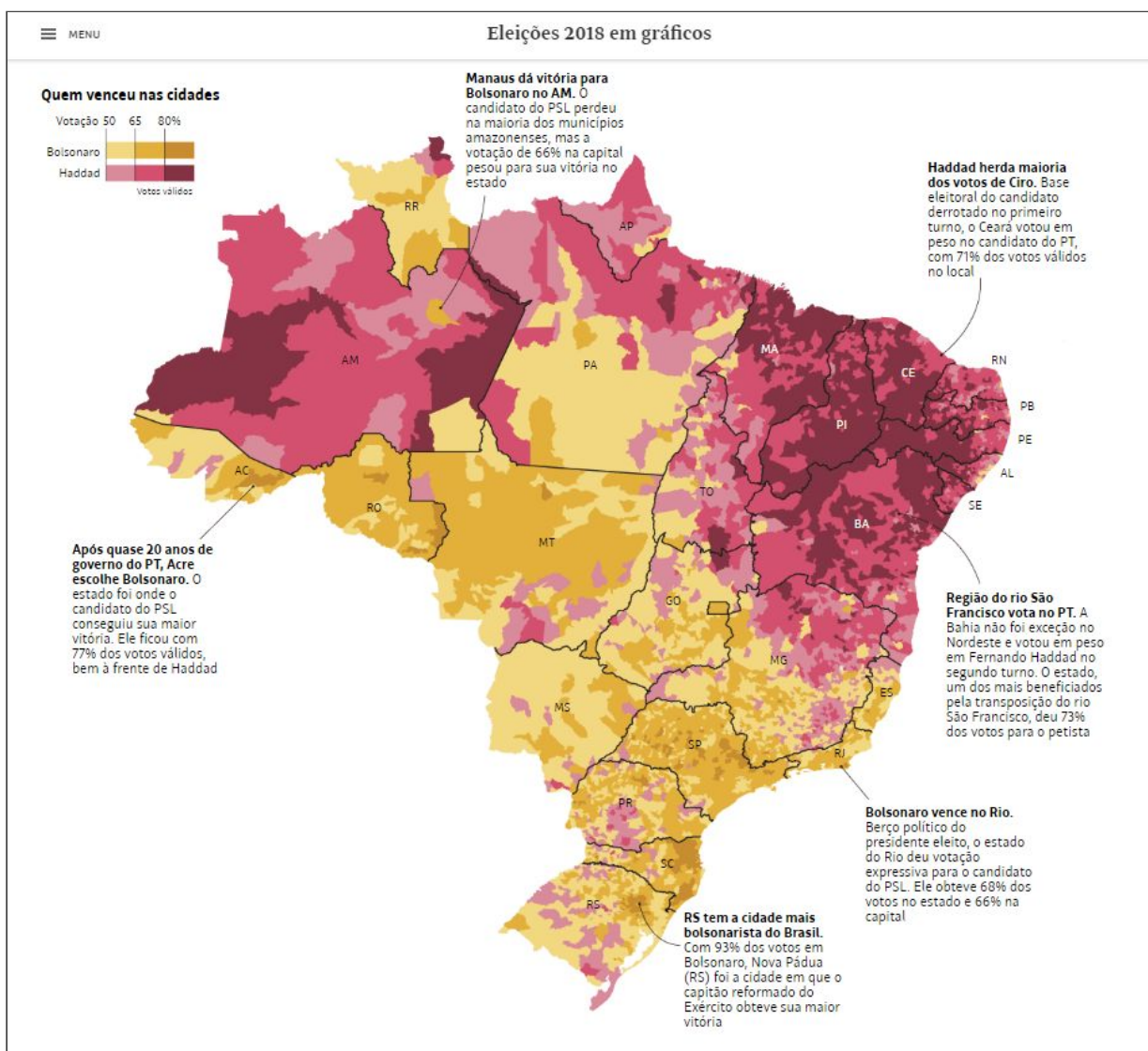
Fonte: O Estado de S. Paulo (2018)

Fundada em 1921, a *Folha de S.Paulo* foi a primeira redação informatizada na América do Sul, com a integração do uso de internet em suas rotinas de apuração. Em 1995, o jornal lançou a *Folha Online*, um dos primeiros portais de notícias em tempo real em língua portuguesa (HISTÓRIA, 2019, online). Dezesete anos depois, a Folha anunciou o FolhaSPDados, um projeto que tinha o objetivo de, com base em dados abertos, reunir mapas e infográficos interativos sobre os temas prioritários relacionados às eleições para a prefeitura da capital paulista (FOLHA, 2012, online).



Esse projeto é fruto da parceria entre a Folha e o programa *Knight International Journalism Fellowship*, do Centro Internacional para Jornalistas (ICFJ). Até então o jornal já vinha trabalhando, ainda que sutilmente, no desenvolvimento e inserção de recursos de visualização de dados em suas matérias, mas nada era estritamente denominado ou relacionado às técnicas de JD. Entretanto, em maio deste ano, o jornal anunciou pelo seu *Twitter* a oficialização do seu Núcleo de Inteligência, que desde 2017 já trabalhava na coleta de dados e na produção de reportagens baseadas nesses insumos, passando a se chamar DeltaFolha, e enfim apresentava-se como a equipe de Jornalismo de Dados da Folha.

Figura 5 - Exemplo de infográfico desenvolvido pela DeltaFolha para o projeto GPS Eleitoral



Fonte: Folha de S. Paulo (2018)

Um dos maiores e mais premiados projetos com dados da DeltaFolha é o GPS Eleitoral, lançado no decorrer das eleições de 2018. Uma ferramenta que executava a captura de postagens dos candidatos nas redes sociais *Facebook* e *Twitter*, assim como transcrevia as falas dos presidentiáveis em vídeos do *YouTube* e na televisão. Tudo isso foi possível pela implementação de uma técnica chamada modelagem de tópicos (*topic modeling*), que resultou na captura de mais de um milhão de palavras ao longo da corrida eleitoral (DATA JOURNALISM AWARDS, 2019, online).

Outras equipes de veículos tradicionais também estão publicando, progressivamente, trabalhos de muita qualidade com dados, e garantindo indicações a premiações internacionais por suas reportagens. Entretanto, veículos menores (nativos do ambiente online e/ou regionais) surgiram com o tempo, já dedicados à produção especializada de notícias e conteúdos especiais guiados por dados. A exemplo disto, podemos citar o jornal *Nexo*, a agência *Volt Data Lab* e o jornal *Metrópoles*.

Lançado em 2015, o *Nexo* é um dos principais nomes do atual jornalismo independente do país. Alinhado ao objetivo de “trazer contexto às notícias e ampliar o acesso a dados e estatísticas” (SOBRE, 2019, online), o jornal digital, em seu curto tempo de atuação, já acumula importantes premiações. Integrando os valores de seus criadores — um cientista político, uma engenheira e um jornalista —, o veículo atua, desde suas primeiras publicações, sob um *workflow* multidisciplinar na produção de reportagens aprofundadas buscando, de maneira visual e moderna, clarificar os dados e informações que estão sendo apresentados em seus conteúdos.

Já a *Volt Data Lab* se apresenta como “uma agência de pesquisas, análises e estudos orientados por dados no setor de mídia e comunicação” (SOBRE, 2019, online). Também atuante no cenário de mídias independentes, a *Volt* disponibiliza para uso livre de seu público os relatórios, dados e visualizações de dados

desenvolvidos por sua equipe, assim como oferece serviços pagos para o desenvolvimento de conteúdos com dados.

O *Metrópoles*, um veículo de notícias do Distrito Federal, lançou em setembro deste ano uma nova editoria, a (M) Dados, um núcleo exclusivo para a produção de Jornalismo de Dados. Segundo artigo divulgado no site do veículo sobre o lançamento, assinado pela equipe da nova editoria — Ricardo Taffner, Rafaela Lima e Lucas Marchesini:

O objetivo do (M)Dados vai além do de produzir notícias baseadas em informações disponíveis em sites públicos, mas fazer análises e cruzamentos entre os dados abertos com a ajuda de tecnologia, buscar conteúdos não divulgados pelos órgãos governamentais por meio da Lei de Acesso à Informação (LAI) e encontrar materiais relevantes nas mais variadas origens existentes. [...] No dia a dia, o núcleo subsidiará as diversas editorias do site para qualificar as matérias publicadas ordinariamente. Ao mesmo tempo, irá se dedicar a reportagens especiais, com um minucioso trabalho de obtenção, raspagem e tratamento de dados a serem apresentados em infográficos dinâmicos e interativos. Os temas não serão voltados apenas ao mundo político, uma vez que serão abordados assuntos variados, como cultura, entretenimento e esportes (TAFFNER, LIMA e MARCHESINI, 2019, online).

Entretanto, talvez pelo caráter muito recente do núcleo, a (M)Dados ainda não trabalhou com visualizações de dados mais modernas ou interativas. O que nos leva a deduzir que, até o momento, em meio às reportagens de dados da editoria, os gráficos e tabelas estáticas publicados são elementos visuais bastante básicos, se comparados ao restante da produção do mercado jornalístico guiado por dados.

Fechamos este capítulo com o destaque de um “serviço” bastante importante para a comunidade mundial de jornalistas de dados e interessados: a disponibilização dos projetos técnicos, códigos-fonte, APIs e outros materiais utilizados na composição das reportagens guiadas por dados. A maioria dos veículos e mídias — inclusive os que citamos anteriormente — têm utilizado a plataforma GitHub<sup>9</sup> para disponibilizar seus insumos.

---

<sup>9</sup> O *GitHub* é uma plataforma para desenvolvedores, focado no armazenamento sob licença de código-aberto de conteúdos e projetos diversos, com a possibilidade de revisão, edição, gerenciamento e compartilhamento de arquivos. Veja mais sobre: <<https://github.com>>

Este movimento está bastante relacionado ao conceito de *open source* (código aberto), pensamento cada vez mais integrado à rotina do jornalista de dados, em paralelo a um outro tipo de jornalismo, de atuação e objetivo comuns ao JD, o Jornalismo Cidadão (TARGINO, 2009) — também chamado de Jornalismo Colaborativo (BELOCHIO, 2009, p. 12; ZANOTTI, 2010, p. 31), ou ainda Jornalismo *Open Source* (MOURA, 2002, online; NOGUEIRA, 2002, online).

O compartilhamento voluntário das equipes e de jornalistas independentes de dados em todo o mundo é um importante fator para a expansão do JD, pois essa troca de conhecimento possibilita que novos projetos, cada vez mais inovadores e completos, sejam produzidos de forma que a informação chegue até o leitor, o público-alvo, em formatos mais refinados, mais interativos e com informações precisas.

## CAPÍTULO 2 - DADOS, INFORMAÇÃO E CONHECIMENTO

Na primeira sessão deste trabalho, apresentamos os princípios básicos e históricos necessários para uma introdução ao entendimento do Jornalismo de Dados. Contudo, antes de ingressar no universo dos principais atores e propagadores desse tipo de jornalismo e dos fluxos e técnicas de produção de reportagens guiadas por dados, é imprescindível compreender o que são esses tais dados, suas particularidades, onde este se relaciona com os conceitos de informação e conhecimento, e como são transformados em insumos com valor noticioso.

Em geral, costuma-se relacionar o termo dado a qualquer tipo de informação existente — um senso comum que não é equivocado. Entretanto, os tipos de dados sobre os quais discorreremos neste capítulo, e sobre os quais interessam ao Jornalismo de Dados, são aqueles existentes no mundo digitalizado e que só podem ser acessados, e/ou manipulados, por intermédio de computadores aparelhados com ferramentas (*softwares*) especializadas para a execução de tarefas como localização, processamento e visualização desses elementos.

Etimologicamente, o substantivo *dado* procede de uma expressão do árabe clássico (*ʿaḏād*), que significa *números*. Em um emprego mais geral, Houaiss (2001) define dado como “aquilo que se conhece e a partir do que se inicia a solução de um problema, a formulação de um juízo, o desenvolvimento de um raciocínio” — definição que se aproxima bastante do senso comum citado anteriormente.

Uma das definições que mais se relacionam com o escopo de estudo deste trabalho foi formulada por Ferreira em 1999, na qual considera-se dado “todo elemento de informação, ou representação de fatos ou instruções, em forma apropriada para armazenamento, processamento ou transmissão por meios automáticos” (FERREIRA, 1999, p. 602 apud CORREIA, 2009, p. 39). Houaiss, neste mesmo raciocínio, e de forma bastante sucinta, afirma que o dado é a

“informação capaz de ser processada por um computador” (HOUAISS, 2001, pg. 903 apud CORREIA, 2009, p. 38).

É importante observar que um dado, por si só, não possui significado relevante e não conduz a nenhuma compreensão. Representa algo que não tem sentido, a princípio, e portanto não tem valor algum para embasar conclusões, respaldar decisões ou sustentar uma pauta e a produção de uma reportagem. Como apontado pelos autores, o tratamento, a ordenação e a organização dos dados é o que reporta à informação.

Acerca dessa relação entre *dado* e *informação*, Donald Knuth (1996) afirma que o primeiro termo se refere à representação do valor, ou quantidade medida, ao passo que *informação* é o significado daquele dado. Enquanto que o *conhecimento* é o resultado de várias informações, organizadas de forma lógica e suficiente para criar um evento, tornar possível um evento ainda não conhecido ou o poder de entender um evento, suas causas, eventos anteriores e suas causas, eventos da causa, evento resultante de causas e o poder de manipular eventos e causas (KNUTH, 1996).

No texto *Relações mútuas entre informação e conhecimento* (2010), de Rodolfo Coutinho Moreira Xavier e Rubenildo Oliveira da Costa, ambos mestres em Ciência da Informação pela PUC-Campinas, a informação e o conhecimento são simultaneamente causa e efeito de si mesmos; a informação é o material direto, a matéria-prima que compõe o conhecimento. Compreende-se aqui um fluxo importante onde o conhecimento depende da informação para ter sentido e esta depende dos dados para se consolidar, e o tratamento e a manipulação desses elementos envolve um processamento cuidadoso para que deles possam ser extraídos insumos assertivos.

## 2.1 Tipos de Dados, *Big Data* e Bases de Dados

Os dados são extratos mínimos de todos os conteúdos, notáveis ou não, que constituem esse cenário digital tão singular que inseriu a humanidade no que conhecemos hoje como a Era da Informação, o período pós-industrial. E no meio de todas estas transformações encontram-se os jornalistas, que tradicionalmente sempre trataram as informações de maneira analógica — coletando dados em formatos físicos, como em documentos impressos, anotações, gravações em rádios etc —, mas agora estão inseridos em um ambiente de dados binários, com imagens técnicas e em um universo mediado, processado e armazenado por máquinas computacionais (LIMA, 2017, p. 1).

Além de serem abundantes, os dados podem ser classificados em diversas categorias, mas, as duas principais são: *dados qualitativos* e *dados quantitativos*. Segundo a Escola de Dados, de forma breve, os *dados qualitativos* tratam de tudo o que se refere à qualidade de algo, como uma descrição de cores, textura, uma descrição de experiências, uma entrevista. Enquanto que os *dados quantitativos* são aqueles que se referem a números, como a quantidade de carros num estacionamento, o tamanho de um objeto, o preço de um produto, a nota em uma prova etc. Além desta distinção acerca do tipo, há também a separação básica, mas de extrema importância para o raciocínio construímos ao longo deste capítulo, entre os dados não estruturados e os dados estruturados.

Quando falamos da troca de informação entre humanos e para humanos, como uma conversa presencial entre dois indivíduos, ou ainda uma troca de mensagens e/ou e-mails, tratamos então de *dados não estruturados*, ou seja, “dados textuais ou não textuais como imagens, cores, sons e formas” (SANTO, 2009) que não têm uma estrutura fixa determinada e assim, mesmo que estes representem informações bem organizadas ao olho humano, não são passíveis de serem lidos ou entendidos por computadores.

Enquanto que os *dados estruturados*, aqueles de natureza numérica, quantificável, repetitiva e previsível — comumente relacionados com planilhas, catálogos de endereço, transações financeiras, desenhos técnicos entre outros —, são “dados dispostos em representações rígidas, sujeitas a regras e restrições impostas pelo esquema que os criou” (ALMEIDA, 2002, p. 5), os quais são armazenados pelos *softwares* que os produzem para que possam ser utilizados em formato binário ou em texto.

Devido a grande e crescente diversidade de dados, surge um termo que enquadra tantos os diferentes tipos de dados como suas origens: o *Big Data*, que trata, de forma sucinta, de gigantescas quantidades de dados estruturados, não-estruturados e também semiestruturados — um tipo híbrido de dado derivado de características dos dois primeiros. O *Big Data* compreende múltiplos conceitos e por isto costuma soar como uma ideia confusa e pouco definida, mas é reconhecido por muitos que a primeira proposta de conceito para o termo foi formulada por Doug Laney, em 2001, ao analisar dificuldades enfrentadas pelas empresas na gestão de dados (MARTINS, SIMÕES E SÁ, 2014, p. 2).

Assim, Laney estruturou três dimensões acerca dos dados: Volume, Velocidade e Variedade. O Volume determina as grandes quantidades de dados, a Velocidade implica na necessidade de captar, armazenar e analisar dados de forma rápida e a Variedade indica a capacidade de tratar, de forma integrada tipos de dados com características distintas (LANEY, 2001, p 1 - 4).

Entretanto, a novidade no conceito de *Big Data* não está atrelada somente à necessidade do tratamento de numerosos dados, nem em fazê-lo de forma muito rápida, mas sim na perspectiva de manusear diferentes tipos de dados. Para Martins, Simões e Sá (2014, p. 2), se os dados a analisar fossem apenas os típicos dados estruturados, por exemplo atributos numéricos ou alfanuméricos, as tecnologias de bases de dados relacionais seriam suficientes para suprir a necessidade desse tipo de tratamento.



“É a dimensão Variedade (necessidade de processar e analisar dados não estruturados) que introduz uma necessidade disruptiva com a informática atual [Kimball & Ross, 2013]. A estes três vetores iniciais (que passaram a ser referidos como “três V’s”) que originalmente definiam o conceito de *Big Data*, foram associadas outras características que contribuíram para a proliferação do mesmo, nomeadamente o crescimento e massificação do uso das redes sociais e dispositivos móveis [Stonebraker, 2012]. ” (MARTINS, SIMÕES E SÁ, 2014, p. 2)

Como ressaltado pelos autores no trecho acima, as redes sociais foram um vetor de extrema importância no crescimento e popularização do *Big Data*. Além das redes sociais, os dados começaram a se originar e acumular, provenientes de diversas fontes internas ou externas, como *streaming* de dados, logs de navegação em sites da internet, dados geoespaciais, textos, imagens entre outros, podendo estes dados ter diferentes tipos de estruturas (HALPER & KRISHNAN, 2013, p. 4).

Exatamente devido ao surgimento dessas novas origens de dados que Michael Stonebraker, em 2012, sugeriu dois novos V’s complementares às ideias de Laney: Veracidade e Valor. A característica Veracidade permite a classificação das fontes de dados de acordo com sua qualidade, considerando a precisão e a atualidade dos dados fornecidos; o Valor refere-se à importância que os dados utilizados terão nas decisões a tomar (STONEBRAKER, 2012 apud MARTINS, SIMÕES E SÁ, 2014, p. 6).

Com o objetivo de explicitar, e com clareza, a dimensão desse universo de dados no qual todos nós estamos inseridos, trazemos para esta argumentação algumas informações pontuais acerca do crescimento e de perspectivas futuras para esse cenário, onde:

- De 2005 a 2020, o universo digital crescerá cerca de 300 vezes, indo de 130 exabytes para 40.000 exabytes ou 40 trilhões de gigabytes. Deste momento até 2020, o universo digital dobrará aproximadamente a cada dois anos (IDC DIGITAL UNIVERSE STUDY, 2012, p. 1, tradução da autora);

- Destas gigantescas quantidades de dados que a humanidade tem produzido, grande parte não está disponível ao público e sequer estão estruturadas de maneira que seja possível compreendê-las, nem mesmo por aqueles que podem acessá-los e manipulá-los. Isso significa que a extração de informações e produção de conhecimentos, que poderiam ser úteis para a sociedade, não acontecem com a agilidade e a eficácia necessárias para lidar com questões sociais e econômicas do século XXI (DADOS ABERTOS CONECTADOS, 2015, p. 16);
- Criamos diariamente cerca de 2,5 quintilhões de bytes de dados. Colocando isto em perspectiva, 90% de todo o amontoado de dados existentes atualmente na Internet foram criados nos últimos dois anos [de 2015 à 2017] - e com novos dispositivos, sensores e tecnologias emergentes, a taxa de crescimento de dados provavelmente acelerará ainda mais (IBM MARKETING CLOUD STUDY, 2017, p. 3, tradução da autora);
- Em 2014 havia 2,4 bilhões de usuários de internet. Esse número cresceu para 3,4 bilhões em 2016, e em 2017 foram adicionados 300 milhões de usuários da Internet. Em junho de 2019, já existiam mais de 4,4 bilhões de usuários de internet. Esse é um aumento de 83% no número de pessoas que usam a internet em apenas cinco anos (KEMP, 2018, online);
- O investimento em hardware de TI, software, serviços, telecomunicações e staff, que poderia ser considerado como “infraestrutura” do universo digital e telecomunicações, irá crescer 40% entre 2012 e 2020. Como resultado, o investimento por gigabyte (GB) durante esse mesmo período cairá de US\$ 2,00 para US\$ 0,20. Entretanto, o investimento em áreas focadas como gestão de armazenamento, segurança, big data, e computação em nuvem, crescerá consideravelmente mais rápido (CREATIVANTE, 2013, online);
- No Brasil, o número de pessoas que usam a internet subiu de 67% para 70% da população de 2017 para 2018, totalizando 126,9 milhões de brasileiros conectados à rede (PESQUISA TIC DOMICÍLIOS - 2018, 2019, p. 7).

Compreendido agora de maneira nítida a grandeza dos dados que produzimos, assim como a expansão do meio digital, uma questão importante atinge esta discussão: como armazenar e organizar essa infinidade de partículas com potencial informacional? Quanto a isto, precisamos então falar das Bases e Bancos de Dados (BDs). A primeira referência ao conceito de Bases de Dados foi realizada por volta dos anos 60 com o termo Banco de Dados — há certa confusão na diferenciação destas duas terminologias no histórico de literaturas especializadas. Cianconi (1987) sugere conceituar bases de dados como "um conjunto de dados inter-relacionados, organizados de forma a permitir recuperação de informações" e banco de dados como "um conjunto de bases de dados".

Para Heemann (1997, p. 2), que considera a literatura em Ciências da Informação em sua afirmação, as bases de dados são "como um arquivo ou um conjunto de arquivos computacionais no qual são armazenados dados, permitindo a recuperação e atualização de informações". Quanto ao objetivo das bases de dados, Albrecht e Ohira (2000, p. 133) afirmam que a finalidade destas é viabilizar informação atualizada, precisa e confiável, assim como também apontam que uma característica fundamental para uma base de dados é, além de armazenar de forma eficiente os dados, fornecer mecanismos suficientemente competentes no quesito "recuperação".

Entretanto, para que seja possível manusear essas bases e retirar destas informações concretas e utilizáveis, seja para fins estratégicos, comerciais e até mesmo jornalísticos, são necessários instrumentos específicos para realizar esse tipo de operação. E para executar essa tarefa temos os Sistemas Gerenciadores de Bancos de Dados (SGBDs), que são programas de computador (*softwares*) especializados em manter as informações e disponibilizá-las quando demandado (DATASUS, 2019, online).

Os SGBDs apareceram no mercado no fim dos anos 60, baseados nos básicos sistemas de arquivos disponíveis na época, que sequer controlavam o acesso às

bases e nem suportavam vários processos simultâneos (TAKAI, ITALIANO e FERREIRA, 2005, p. 6). Com o tempo, esses sistemas deixaram de ser apenas aplicações de computador e se transformaram em componentes centrais do ambiente computacional moderno (KORTH, HENRY e SUDARSHAN, 2012, p. 1), e hoje são ferramentas indispensáveis para gerenciar informações (RAMAKRISHNAN e GEHRKE, 2011, p. 21).

Mas qual é o impacto desse universo de dados, bases e sistemas no jornalismo? Tom Koch (1991 apud BARBOSA, 2007b, p. 1317), ao avaliar o potencial das bases de dados para o jornalismo, afirmou que a melhor maneira de o descrever seria considerar que as BDs trariam os benefícios da revolução da imprensa do século XVIII para o jornalismo do século XXI, justamente pelo impacto e o nível de mudança que trariam. Esse argumento é evidenciado pela expansão das técnicas de pesquisa e levantamento de informações que propiciaram novas oportunidades de apuração para os jornalistas, como a utilização do computador para analisar grandes volumes de registros governamentais.

Como exemplos históricos desse tipo de técnica, ainda em sua essência mais básica, temos Clarence Jones, do *The Miami Herald*, que em 1969 trabalhou com um computador para encontrar padrões no sistema de justiça criminal; David Burnham, do *The New York Times*, que em 1972 usou ferramentas computacionais para investigar discrepâncias nas taxas de criminalidade relatadas pela polícia; Elliot Jaspin, do *The Providence Journal*, que, em 1986, comparou bancos de dados para expor os motoristas de ônibus escolares com histórico de condução ruim e registros criminais — reportagem esta que inspirou os repórteres da época a utilizarem bancos de dados para investigar instituições e processos do governo.

Esse movimento e as histórias que emergiram dele influenciaram em mudanças na lei e conscientização pública sobre questões de corrupção. Importante citar também Bill Dedman, do *The Atlanta Journal-Constitution*, que recebeu o Prêmio Pulitzer por sua pesquisa em 1988, *The Color of Money*, que tratava da

discriminação de empréstimos hipotecários em bairros com maior parte da população negra e de renda média.

Assim, desde os anos 90 até os dias atuais, as bases de dados continuam sendo uma das principais matérias-primas do jornalismo, especificamente do Jornalismo de Dados. Mesmo com o desenvolvimento das técnicas e ferramentas, e principalmente com a transformação do perfil do jornalista de dados, as BDs ainda são as principais fontes de informação desse tipo de jornalismo. Entretanto, naquela época, o acesso a dados de interesse público demandava enorme esforço de reportagem, pois era necessário convencer autoridades e instituições a concederem essas informações, mas, com o espalhamento das práticas de RAC pelas redações do mundo, pouco a pouco, barreiras acerca da disponibilização de dados com teor público foram sendo ultrapassadas.

No Brasil, essas mudanças se desencadearam graças a uma série de reportagens publicadas no *Jornal do Brasil*, por Mário Rosa, em agosto de 1991. Nas matérias, Rosa denunciava o escândalo do desvio de verbas da Legião Brasileira de Assistência (LBA) por Rosane Collor, esposa do então presidente Fernando Collor de Melo — foi a primeira vez que um jornalista usou dados do Sistema Integrado de Administração Financeira do Governo Federal (Siafi), criado em 1987, que registra os pagamentos feitos pelo governo federal (GRANDIN, 2014a, p. 3).

Anos depois a *Folha de S. Paulo*, com os jornalistas Fernando Rodrigues e José Roberto Toledo à frente dos movimentos de mudança no jornal, foi que os conceitos e técnicas de produção de reportagens assistidas por computador se disseminaram na redação. Um dos resultados dessas transformações foi o desenvolvimento do banco de dados Políticos do Brasil, idealizado por Rodrigues em 1998 —, porém, nos anos 2000, quando o jornalista fundou o Poder360, o banco de dados foi transferido para o domínio do jornal digital onde continua sendo atualizado até hoje.

Grandin (2014a, p. 3) também destaca como importantes para esse momento de ações na defesa da facilitação do acesso à informação, a fundação da Associação Brasileira de Jornalismo Investigativo (Abraji), em 2002, e os papéis das ONGs Contas Abertas e Transparência Brasil, que foram significativos atores na divulgação de bases de dados de interesse público no país.

Na esteira do avanço da web, surgiram também outras mobilizações em defesa da divulgação de informações de interesse público por governos e instituições privadas, alinhadas com as características da internet – aberta, não-hierárquica, descentralizada. Governos, empresas e instituições passaram então a produzir e lançar uma torrente de bases de dados na internet, num volume que cresce exponencialmente a cada ano (GRANDIN, 2014a, p. 3).

Todavia, mesmo com a difusão voluntária dessas informações ao longo dos anos 90, essas práticas só vieram a se consolidar, em nível mundial e sistemático, somente na década seguinte, em virtude do aparecimento e fortalecimento de movimentos pela transparência governamental, pela liberação de informações de interesse do público e pelos dados abertos.

## 2.2 Dados Abertos

Ao tratar de Jornalismo de Dados, é imprescindível falar sobre dados abertos (*open data*) e o desenvolvimento das leis que instituíram o acesso à informação. Em comparação com a legislação norte-americana — que colocou em voga, ainda em 1966, a *Freedom of Information Act* (FOI) —, o Brasil processou tardiamente a ideia de assegurar em lei o acesso livre aos dados do governo e foi somente em novembro de 2011 que a Lei de Acesso à Informação (LAI), ou Lei nº 12.527, foi promulgada no país no âmbito do direito à informação. Naquele mesmo ano o governo brasileiro foi um dos fundadores da *Open Government Partnership* (OGP), ou Parceria para Governo Aberto, que conta atualmente com a participação de 65 países.

A instituição da LAI resultou na criação do Portal Brasileiro de Dados Abertos, em 2012, assim como na edição da Política de Dados Abertos do Poder Executivo Federal, com a publicação do Decreto nº 8.777/2016 que, além de tratar sobre os

objetivos e funcionamento da nova política, também incluiu à LAI a possibilidade de solicitação da abertura de bases de dados pelos mesmos canais de transparência passiva já regulados pela lei (POSSAMAI e SOUZA, 2017, p. 6). Importante dizer que, quando falamos de dados abertos, não tratamos apenas de dados públicos de origem governamental, pois informações também podem ser abertas voluntariamente por organizações privadas, em ações que beneficiem o interesse público.

O movimento *open data* teve sua ascensão e propagação efetiva na última década, mas o termo “dados abertos” foi utilizado pela primeira vez em 1995, em um documento de uma agência científica americana que tratava da divulgação de dados ambientais e geofísicos, no qual os autores desenvolveram uma troca de informações científicas entre diferentes países, com o objetivo de analisar e compreender fenômenos naturais globais — “o princípio de bens comuns aplicado ao conhecimento já foi teorizado por Robert King Merton, em 1942, quando sua teoria mostrou os benefícios de dados científicos abertos” (CHIGNARD, 2013 apud ISOTANI e BITTENCOURT, 2015, p.21).

Desde então, o *open data* tem sido fortalecido por movimentos mundiais como, em 2007, quando onde ativistas da internet se uniram para definir um conceito para dados abertos públicos, ou governamentais. A premissa resultante do debate foi que esses dados são “propriedades comuns, da mesma forma que as ideias científicas”; um conceito inspirado nas formulações sobre código aberto, ou *open source* (CHIGNARD, 2013 apud ISOTANI e BITTENCOURT, 2015, p.21).

O Portal Brasileiro de Dados Abertos — que adota a definição da *Open Definition* (2014) — descreve dados abertos como sendo os dados que qualquer pessoa pode acessar, utilizar, modificar e compartilhar com qualquer finalidade, desde que preservem sua proveniência (sua fonte original) e sua abertura (compartilhamento pelas mesmas licenças que foram obtidas). Essa definição traz consigo três normas fundamentais, também conhecidas como as três leis dos dados

abertos, que foram propostas por David Eaves, especialista em políticas públicas e ativista do movimento *open data*, e são sustentadas pela *Open Knowledge Foundation*. Estas três propriedades são:

- **DISPONIBILIDADE E ACESSO:** os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, e preferencialmente devem ser possíveis de ser baixados pela Internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável;
- **REÚSO E REDISTRIBUIÇÃO:** os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados;
- **PARTICIPAÇÃO UNIVERSAL:** todos devem ser capazes de usar, reutilizar e redistribuir – não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso “não comercial” que impediriam o uso “comercial”, ou restrições de uso para certos fins (ex.: somente educativos) excluem determinados dados do conceito de “abertos” (ISOTANI e BITTENCOURT, 2015, p. 20).

Além do estabelecimento de um conceito e do entendimento das três principais características, outro aspecto importante acerca dos dados abertos é a sua fonte, a origem deles. Em suma, esse tipo de dado pode ser originário de qualquer fonte, desde que seja do interesse público a sua disponibilização e que as informações provenientes destes dados não tenham caráter confidencial e/ou que possam trazer riscos à segurança de pessoas ou instituições.

Para o *Open Data Handbook* (2019), todos os dados carregam em si o potencial de serem abertos, principalmente dados do governo, de empresas e de instituições de ensino e pesquisa. O ponto crucial na discussão do que pode ou não ser um dado aberto é que o foco da disponibilização desse dado não pode estar relacionado a dados pessoais, ou seja, dados que contenham informações sobre indivíduos específicos (OPEN DATA HANDBOOK, 2019, tradução da autora).

Segundo a *Open Knowledge Foundation* (2014 apud ROCHA, 2014, p. 22) existem alguns tipos de dados abertos que são detentores de grande potencial de uso e aplicação. Esses tipos estão ligados aos setores de cultura, ciência, finanças, estatísticas, tempo, ambiente e transporte. Todas essas áreas estão ligadas ao



objetivo primário mais formal da transparência e do *accountability* (prestação de contas) mas, quando falamos da ciência, é importante considerar um objetivo secundário crucial na abertura de seus dados: a propagação de informação com o intuito de gerar conhecimento, de propiciar um intercâmbio de ideias e descobertas.

E no meio desse fluxo ininterrupto de informações providas de diversas fontes, figurando ao lado de pesquisadores e ativistas dos movimentos *open data* e *open source*, estão os jornalistas.

Jornalistas de investigação e de dados têm tradicionalmente coletado, utilizado e até mesmo aberto dados na América Latina [e no mundo] enfrentando níveis muito diferentes de acesso a dados públicos. Mesmo em contextos onde leis de acesso à informação não existem ou o fornecimento de dados é imposto em formatos abertos, os projetos reuniram fontes de dados e tornaram esses dados publicamente disponíveis e utilizáveis (CASANUEVA, 2015, online).

O jornalista Juan Manuel Casanueva (2015), em um artigo para o site da IJNET, trouxe um panorama geral da utilização e impactos dos dados abertos no jornalismo, especificamente nos países latino americanos. Casanueva afirma que projetos jornalísticos com foco na abertura de dados, além de produzir grandes reportagens de impacto, podem cooperar para o próprio fornecimento de dados de interesse sobre as temáticas mais básicas no contexto social dos países onde atuam. Ou seja, além de usuários da transparência passiva, os repórteres têm atuado também como desbravadores de dados públicos não divulgados.

Com o aumento crescente de profissionais do jornalismo de dados nessas regiões, o papel do jornalista tem se apresentado cada vez mais influente devido à sua capacidade na análise de cenários, na descoberta de atividades ilegais e no *storytelling* baseado em dados (CASANUEVA, 2015, online).

Se os funcionários do governo trabalhassem mais de perto com os jornalistas de dados, eles teriam uma melhor visão sobre a demanda de dados pública aumentando seu foco para os esforços de abertura de dados e receberiam feedback mais detalhado sobre a qualidade dos dados públicos abertos (CASANUEVA, 2015, online).

A resistência à liberação de dados de interesse público ainda é um dos desafios do Jornalismo de Dados, apesar do crescimento do movimento e das legislações que regulam a abertura das informações públicas no Brasil, e no restante do mundo. Assim como Casanueva (2015), Mazzote (2016) defende que o apoio à defesa da transparência deve estar integrado no cotidiano dos profissionais que lidam com essas massas de dados, com o objetivo de pressionar as instituições responsáveis pelas bases de dados a manterem-nas estruturadas, atualizadas e acessíveis.

Por fim, o JD, ou JGD, apresenta-se então como uma solução; uma prática que pretende transformar essa massa de dados estruturados, abertos ou não, em informação relevante e passível de compreensão por qualquer indivíduo interessado. Neste contexto, as técnicas de programação, ou a integração com programadores, irrompem como “algo primordial para a expressão jornalística, atrelada ao conhecido faro de repórter, à sola de sapato e à vontade de navegar nas histórias ainda inexploradas em mares de dados” (MAZZOTE, 2016, online).

## CAPÍTULO 3 - FLUXO DE PRODUÇÃO E AS FERRAMENTAS

Como descrito no primeiro capítulo desta monografia, o Jornalismo de Dados (JD) é resultante de uma conjunção das práticas de outros três tipos de jornalismo: o Jornalismo de Precisão, a Reportagem Assistida por Computador e o Jornalismo Investigativo. Essa mescla de vertentes se reflete fortemente na estruturação dos processos produtivos da reportagem baseada em dados, que também sofre influência das premissas mais básicas do jornalismo tradicional, como a pirâmide invertida e as cinco perguntas — técnicas que João Canavilhas (2006, p. 5) relaciona da seguinte forma: “a redação de uma notícia começa pelos dados mais importantes; a resposta às perguntas o quê, quem, onde, como, quando e por quê, seguido de informações complementares organizadas em blocos decrescentes de interesse”.

### 3.1 Sobre os Processos

Ainda na discussão acerca da sua conceituação, os processos de trabalho do JD já são colocados como ponto importante na definição deste tipo de jornalismo. Há autores que se baseiam exatamente nos fluxos e técnicas do Jornalismo de Dados para delimitarem a conceituação dessa prática — como Meyer, Lorenz, Bradshaw, Rogers e Silver. Uma das formulações mais antigas sobre processos para divulgação jornalística de dados é de Philip Meyer, em 1991, quando ainda escrevia sobre Jornalismo de Precisão, um dos precursores do Jornalismo de Dados que conhecemos hoje. Na obra intitulada *The New Precision Journalism*, Meyer (1991 apud LIMA, 2017, p. 4 e 5) listou seis recomendações fundamentais para uma comunicação adequada de dados.

1. Reúna-o. Querendo ou não você nunca tentará imitar cientistas em seus métodos de coleta de dados, você pode lucrar se souber alguns de seus truques. É sempre bom lembrar, como o professor H. Douglas Price disse-me em Harvard na primavera de 1967, que "os dados não vêm da cegonha."
2. Armazene-o. Jornalistas à antiga armazenam dados em pilhas de papel em suas mesas, em cantos de seus escritórios, e, se eles são realmente bem organizados, em grampo-arquivos. Computadores são melhores.
3. Recupere-o. As ferramentas do jornalismo de precisão podem ajudá-lo a recuperar dados que você mesmo recolheu e armazenou, dados que

alguém armazenou, ou ainda dados que alguém armazenou por motivos completamente alheios a seu interesse.

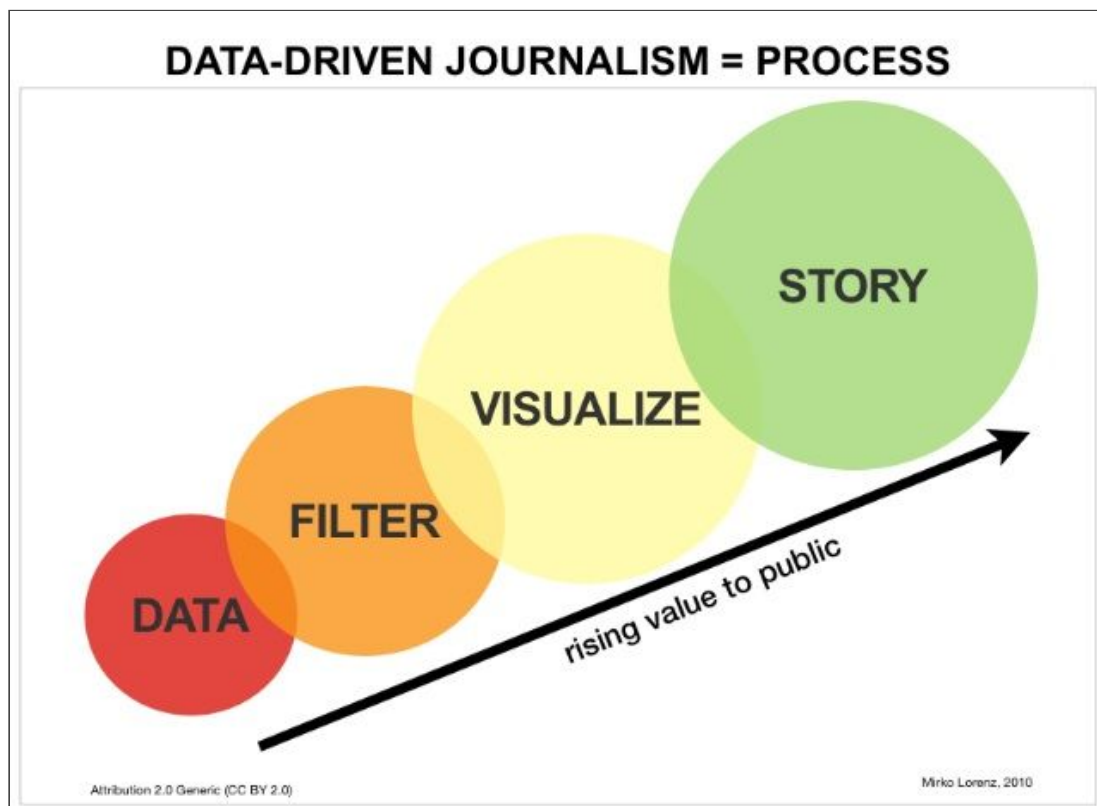
4. Analise-o. Análise jornalística muitas vezes consiste em apenas triagem para encontrar e listar os desvios interessantes, mas também pode envolver pesquisas para o nexos de causalidade implícita, para os padrões que sugerem que fenômenos diferentes variam juntos por razões interessantes.

5. Reduza-o. Redução de dados tornou-se tão importante no jornalismo como a coleta de dados. Uma boa notícia é definida pelo que deixa de fora, bem como o que inclui.

6. Comunique-o. Um relatório não lido ou não entendido é um relatório desperdiçado. (MEYER, 1991, p. 33 apud LIMA, 2017, p. 4 e 5)

Aproximadamente duas décadas depois, Mirko Lorenz (2010, online), em uma apresentação para a mesa redonda organizada pelo *European Journalism Centre* (EJC) sobre Jornalismo Guiado por Dados, mostrou um diagrama com o que seriam as novas fases do fluxo de trabalho do JGD (ou JD).

Figura 6 - Fluxo de trabalho de Lorenz para Jornalismo de Dados



Fonte: Lorenz (2010)

O esquema de Lorenz determinou quatro etapas para o processo produtivo do JD, que seriam: *data* (a coleta dos dados), *filter* (a filtragem dos dados), *visualize* (a

visualização dos dados) e *story* (a construção da notícia baseada nos dados). A estrutura bastante simples apresentada pelo autor considera um aspecto importante no fluxo de trabalho jornalístico com dados ao determinar a indicação *value to public*. Com isso é possível entender que o avanço no processo de refinamento, de apuração dos dados, feito pelo jornalista, aumenta o valor que a história a ser contada terá para o público. Ou seja, até que o jornalista/repórter faça o trabalho de traduzir aquelas partículas com potencial informacional para um formato inteligível — nesse caso, a reportagem —, os dados por si mesmos possuem mínimo valor para o público.

Paul Bradshaw (2011, online) trata da estrutura do fluxo de trabalho do JD com uma abordagem baseada no conceito clássico de pirâmide invertida. Nesta estrutura de Bradshaw estão contempladas as etapas de um produto do Jornalismo de Dados. O diagrama de pirâmide é dividido em: *compile* (compilar), *clean* (limpar), *context* (contextualizar) e *combine* (combinar). Como uma extensão desse primeiro processo de quatro fases, o autor incorpora no seu *workflow* o momento *communicate*, que trata da comunicação dos dados para o público. Esse momento gera uma nova pirâmide<sup>10</sup> que descreve como se deve comunicar os resultados de uma apuração com dados, que considera a visualização, a narração, a socialização, a humanização, a personalização e a utilização.

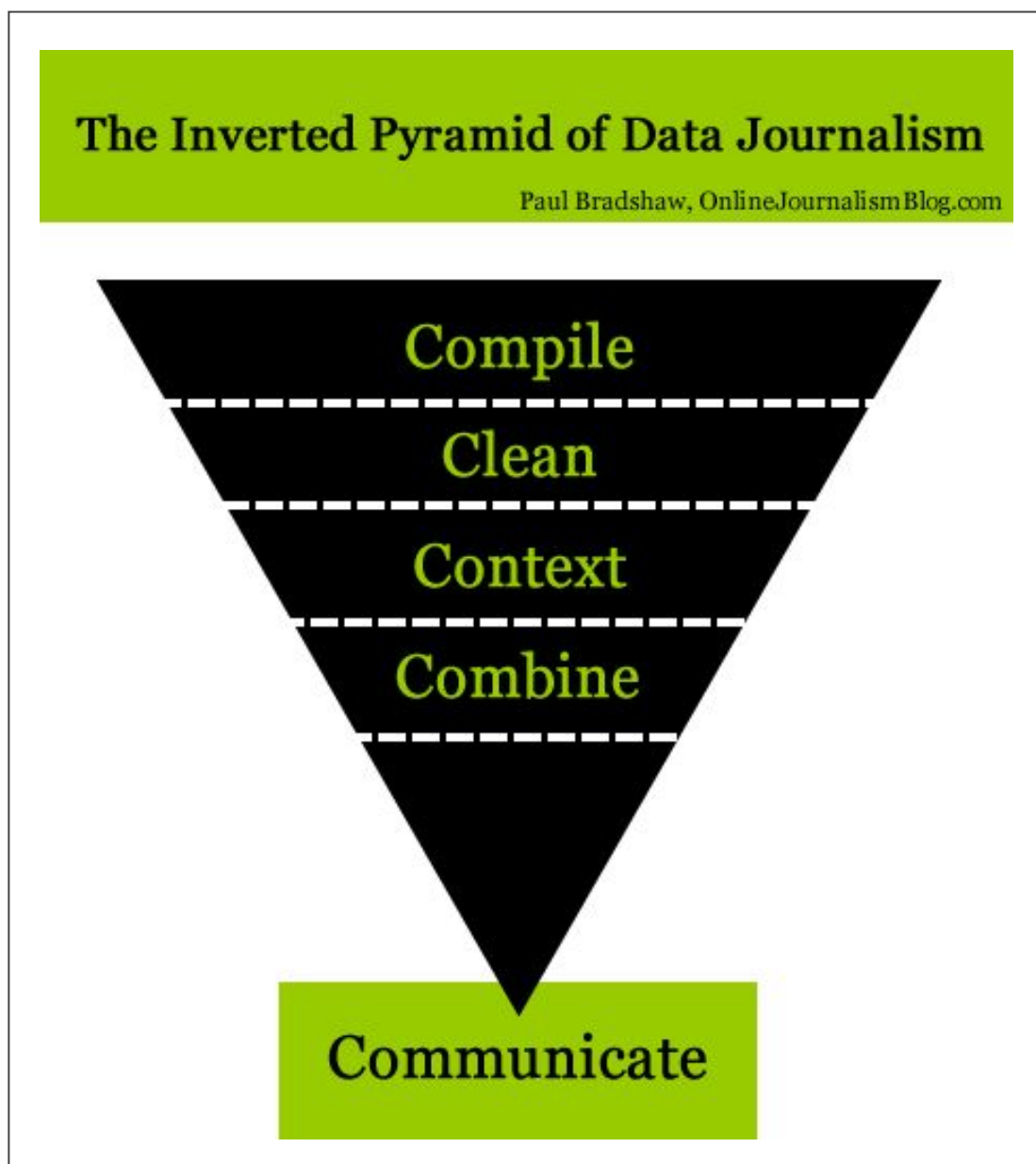
Distintamente dos autores anteriores, Simon Rogers (2013, online) apresenta uma estrutura mais complexa para o trabalho com Jornalismo de Dados. Em seu artigo *A data journalism workflow*, Rogers explica que antes de um conjunto de dados se transformar em uma história jornalística, há todo um processo de seleção, refinamento e categorização desses dados. Esse processo está dividido em “aproximadamente 70% arrumando os dados e 30% fazendo as coisas divertidas de visualizá-los e apresentá-los” (ROGERS, 2013, online, tradução da autora). Estes 70% do processo são fracionados nos seguintes estágios: *share data* (análise das

---

<sup>10</sup>Veja mais sobre a segunda parte da formulação da pirâmide invertida de Bradshaw no link: <https://onlinejournalismblog.com/2011/07/13/the-inverted-pyramid-of-data-journalism-part-2-6-ways-of-communicating-data-journalism/>

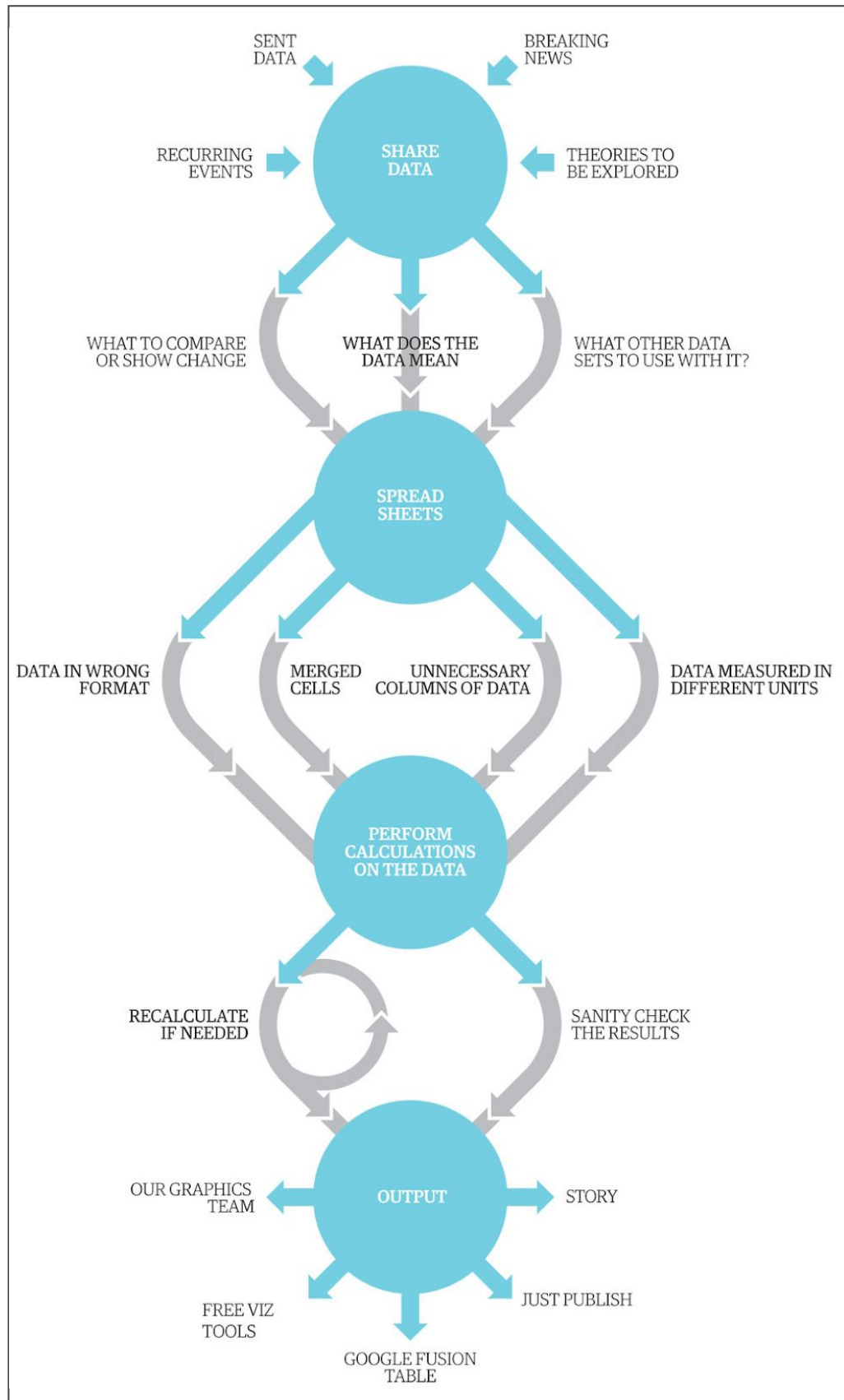
possibilidade de trabalho com os dados, combinação com outros conjuntos de dados, demonstração de alterações ao longo do tempo), *spread sheets* (organização sistemática das planilhas com os dados), *perform calculations on the data* (realização dos cálculos dirão se há uma história ou não e verificação da veracidade das informações), e enfim *output* (o formato de saída das informações consolidadas, podendo ser como reportagem, infográfico ou outros tipos de visualização).

Figura 7 - Pirâmide Invertida de Bradshaw



Fonte: Bradshaw (2011)

Figura 8 - Workflow de Rogers para JD



Fonte: Rogers (2013)

Rogers (2014, online) defende que, no processo de produção de reportagens de dados, os repórteres ainda precisam fazer as cinco perguntas básicas do jornalismo (quem, o quê, quando, onde e por quê). Segundo o autor, no contexto do Jornalismo de Dados, as cinco perguntas se aplicariam da seguinte forma:

Tabela 1 - Cinco Porquês para Jornalismo de Dados

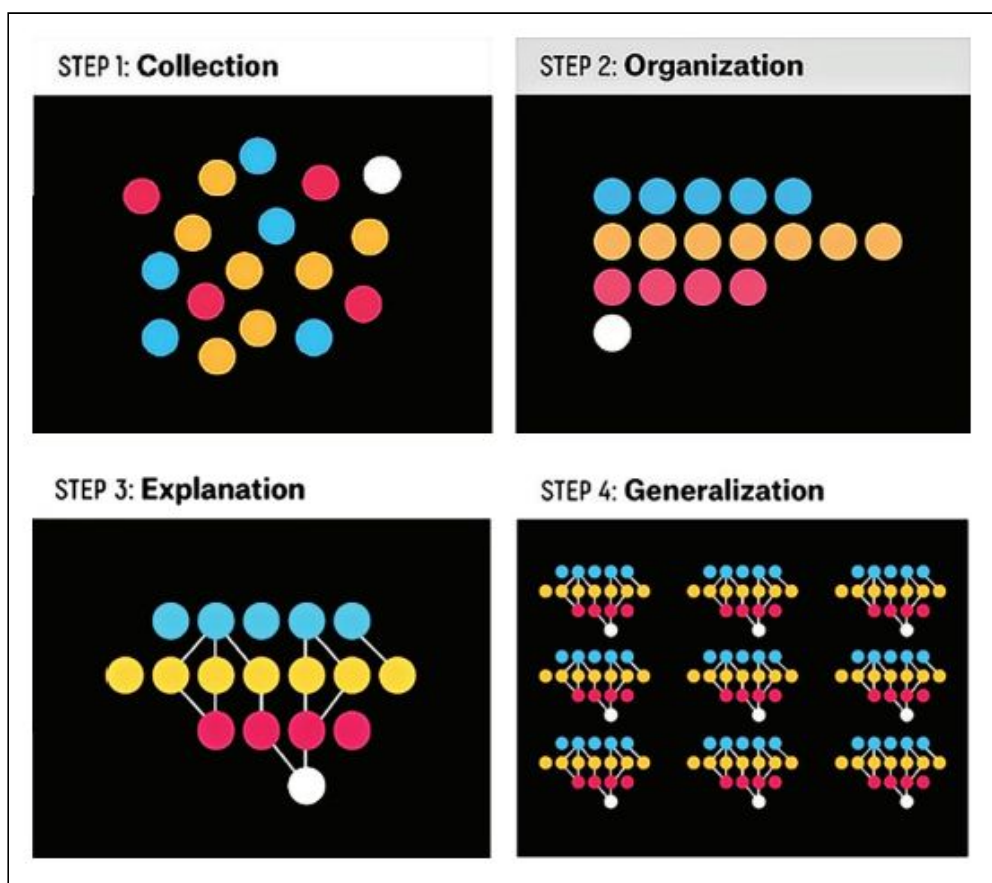
<b>QUEM?</b>	Qual a origem dos dados? Eles são confiáveis?
<b>O QUÊ?</b>	Qual o significado desses dados? Qual informação pode ser retirada deles? Qual o nível de interesse público desses dados?
<b>QUANDO?</b>	Quão antigos são os dados? Estão ultrapassados? Qual a taxa de atualização?
<b>ONDE?</b>	Qual a geolocalização dos dados? Eles se referem ao mesmo local? Tratam de informações nacionais ou de extratos específicos locais?
<b>POR QUÊ?</b>	O que esses dados mostram? Qual é o tamanho de algo? Está aumentando ou diminuindo? Como se compara a outra coisa/em outro lugar? Existe uma relação de causa e efeito?

Fonte: Elaborado pela autora (2019); com base em Rogers (2014)

Outro teórico importante na delimitação dos fluxos para JD é Nate Silver (2014, online), que em seu artigo *What the Fox Knows*, determinou quatro momentos pelos quais os dados são submetidos para que enfim se tornem informação com valor noticioso: *collection* (coleta), *organization* (organização), *explanation* (explicação) e *generalization* (generalização).



Figura 9 - Fluxo do Jornalismo de Dados de Silver



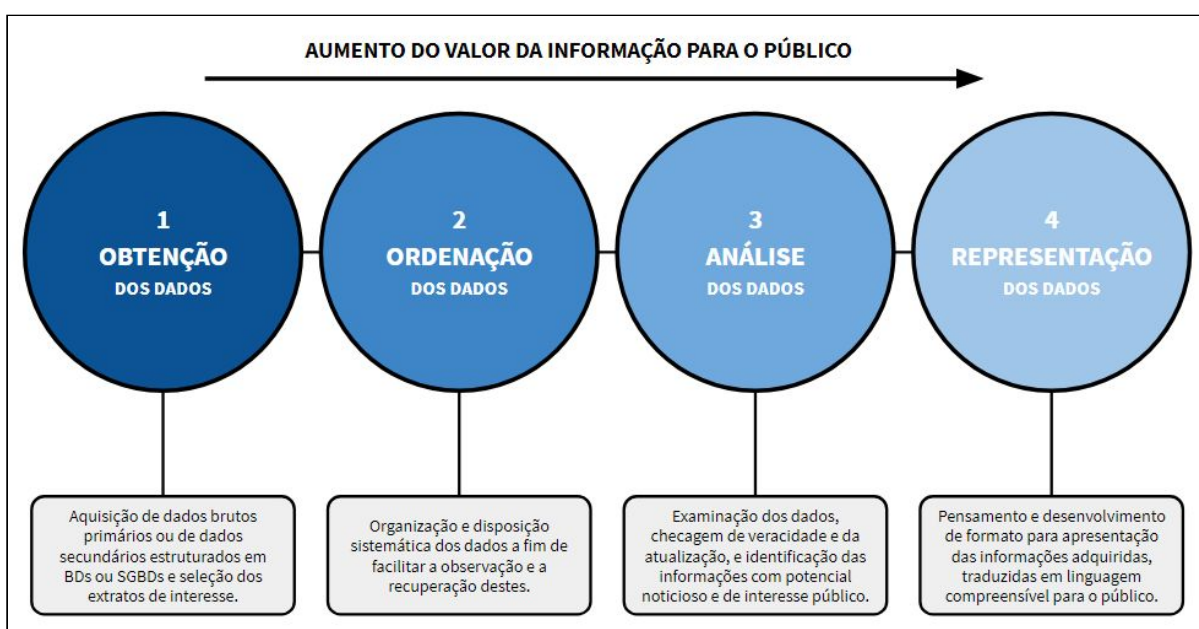
Fonte: Silver (2014)

A primeira etapa, a coleta dos dados, pode ser realizada por meio de questionários, experimentos ou raspagem de dados de sites da internet. A próxima etapa é a organização, que pode ser feita com a execução de estatísticas descritivas, estruturação de um banco de dados ou construção de visualizações a partir desses dados. A explicação, ou explicação, seria a terceira etapa desse processo, onde, além de responder às cinco perguntas básicas, o jornalista de dados executa testes estatísticos para localizar padrões, relações e *outliers*, além de observar aspectos relacionados a causa e efeito.

Por fim, a quarta e última etapa seria a generalização, técnica proveniente do método científico onde verifica-se hipóteses por meio de previsões ou experimentos repetidos, ou seja, um procedimento para inferir sobre o comportamento dos eventos no futuro (SILVER et al., 2014, tradução da autora).

Observando estas cinco principais perspectivas acerca do *workflow* do Jornalismo de Dados — de Meyer (1991), Lorenz (2010), Bradshaw (2011), Rogers (2013) e Silver (2014) —, pode-se perceber que elas são complementares e coincidem em pelo menos quatro etapas fundamentais para a consolidação de uma reportagem guiada por dados, sendo estas a obtenção, a ordenação, a análise e a representação dos dados. Com base nessa constatação, sugerimos então um modelo objetivo para os processos de trabalho em Jornalismo de Dados.

Figura 10 – Sugestão de Fluxo de Produção para Jornalismo de Dados



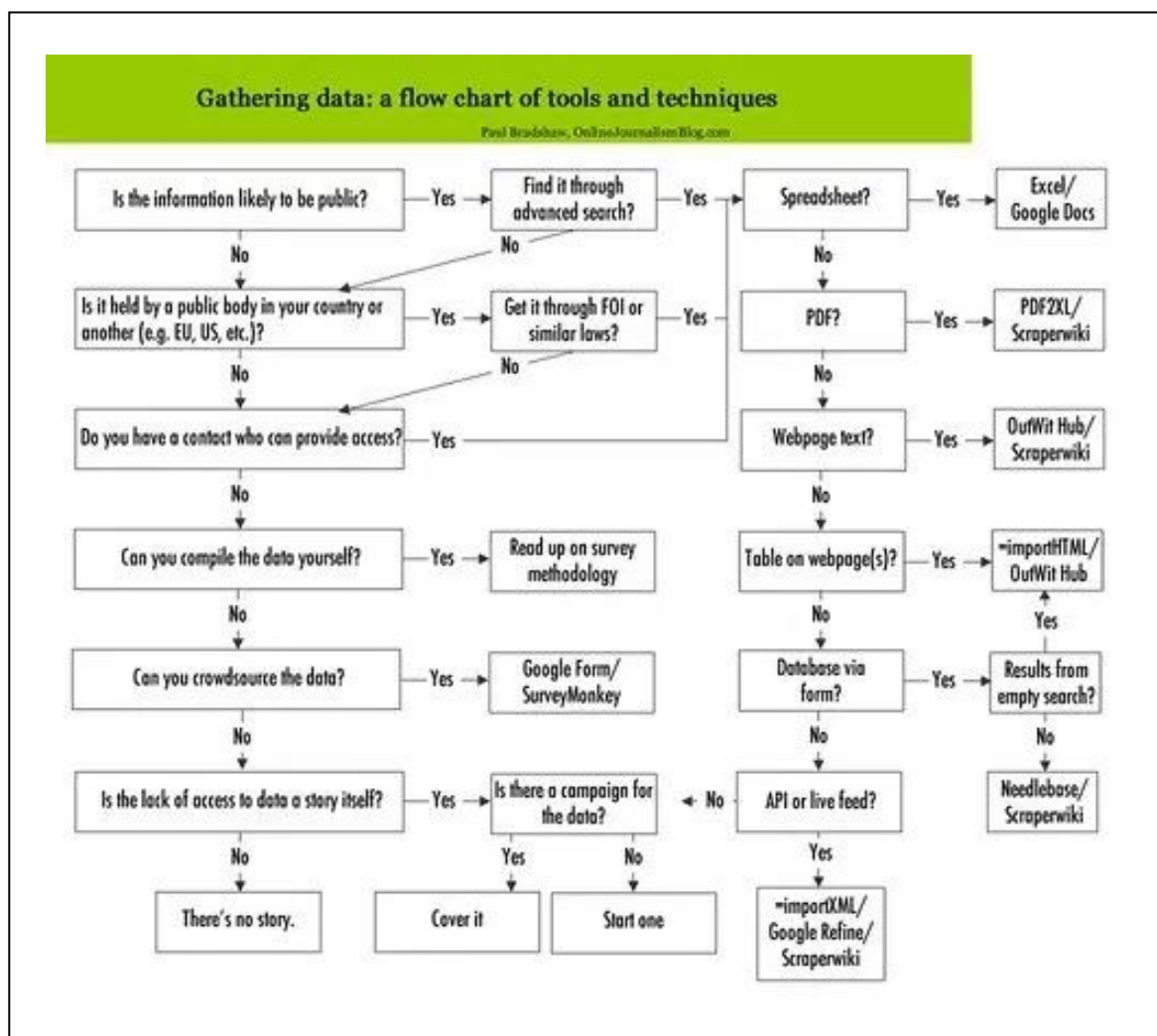
Fonte: Elaborado pela autora (2019)

### 3.2 Sobre as Ferramentas e Habilidades

As ferramentas e habilidades são elementos fundamentais para os fluxos de produção do Jornalismo de Dados. Para além da utilização do computador (*hardware*), o trabalho com dados, de maneira geral, exige o domínio de ferramentas específicas (*softwares*) destinadas ao manuseio e análise desses materiais, assim como habilidades multidisciplinares, que vão além das esperadas de um jornalista com formação tradicional.

Não que os jornalistas devam se tornar programadores, designers ou estatísticos, mas é imprescindível entender o funcionamento, a dinâmica e, principalmente, os potenciais desse novo suporte, para saber como usá-lo. Tradicionalmente, foi possível atuar nos meios de comunicação sem domínio dos números, mas a sobrevivência desses profissionais está cada dia mais ameaçada pelo avanço das bases de dados no dia a dia dos cidadãos (GRANDIN, 2014b, p. 38).

Figura 11 - Etapa de coleta de dados de Bradshaw



Fonte: Bradshaw (2011)

Bradshaw (2011), que citamos anteriormente ao tratar dos fluxos, desenvolveu também um esquema para explicar a complexidade da primeira etapa do seu workflow para JD, a coleta (*compile*). Observando a figura 11 percebe-se facilmente a quantidade de ferramentas e habilidades envolvidas somente na fase inicial do processo de produção baseada em dados.

Esse esquema é uma forte evidência da característica de multidisciplinaridade inerente a um jornalista de dados. Gradin (2014b, p. 38), ao analisar o diagrama de Bradshaw, destaca como requisitos prévios conhecimentos nas áreas do Direito (leis de acesso à informação), Computação (linguagens de programação, *scrapping*) e da Estatística (compilação e análise dos dados). Em sua tese, Grandin também acrescenta à lista o entendimento, ou ao menos a familiaridade, com os campos do Design (formatos e visualização dos dados) e de Narrativas para Novas Mídias.

A gama de opções de ferramentas para jornalistas de dados vem crescendo cada dia mais, acompanhando o surgimento quase diário de novas tecnologias informacionais (plataformas, mídias e equipamentos). Esse movimento não tem sinalizado que irá diminuir seus avanços e isso aponta para a necessidade de atualização constante dos conhecimentos, como uma exigência deste novo cenário para o campo da Comunicação e seus profissionais.

O Jornalismo de Dados, que antes era revolucionário pela manipulação estratégica de planilhas e documentos, hoje se aproxima até mesmo de técnicas de *machine learning* na busca de otimização da apuração, da checagem das informações e de visualizações ou formatos cada vez mais inovadores, que informem e entretenha o leitor em quaisquer suportes, seja no jornal impresso, em um telejornal, no ambiente online, ou, cada vez mais, nas plataformas mobile.

Considerando esse contexto, é possível observar que a escolha de um conjunto de ferramentas para a construção do storytelling no Jornalismo de Dados envolve diversos fatores inerentes ao jornalista/repórter e à estrutura na qual ele está inserido, como por exemplo, a possibilidade de investimento na compra de licenças de programas ou aplicações (*softwares*) mais completas, assim como a aquisição de máquinas computacionais com capacidade de trabalhar com grandes quantidades de arquivos e com diversos formatos de dados. Vale observar ainda que se há o suporte de uma equipe especializada e dedicada à execução de reportagens com dados, além de condições mais individuais ao profissional, como o

grau de conhecimento que ele detém de áreas transversais — Direito, Computação, Estatística, Design e outras —, assim como possíveis limitações em atuar diretamente com linguagens de programação e/ou com aplicações complexas. E, como observado nos modelos apresentados de workflow para JD, são diversas as etapas no processo, e para cada uma destas, pode-se empregar uma ferramenta ou técnica diferente.

### 3.2.1 Google e Ferramentas para Jornalismo de Dados

A empresa Google, por meio da *Google News Initiative*, tem atuado nos últimos anos como um forte apoiador do Jornalismo de Dados no mundo, como citado anteriormente neste trabalho. Além dos programas e parcerias, a empresa contribui para o futuro do jornalismo também com o desenvolvimento e fornecimento gratuito de ferramentas e cursos com foco em dados, novas mídias, universo digital e *fact-checking* (checagem de fatos).<sup>11</sup>

Oferecido pela iniciativa, o curso *Data Journalism* ensina de forma direta, mas bastante completa, como utilizar as ferramentas gratuitas do Google na produção de uma reportagem baseada em dados. No curso são apresentadas as seguintes ferramentas: *Google Sheets*, *Data Studio*, *Google Data GIF Maker*, *Tilegrams*, *Flourish*, *Google Crisis Map*, *Public Data Explorer*, *Google Trends* e o *Google Surveys*.

A primeira aplicação apresentada no curso é a *Google Sheets*, ou Planilhas Google; o recurso mais básico quando falamos de manipulação de dados. Bastante semelhante ao tradicional Microsoft Excel, o *Sheets* permite o trabalho com tabelas e planilhas, incluindo o diferencial de ser uma aplicação online disponível na nuvem, no Drive da Google, passível de compartilhamento e operação simultânea — característica importante para apurações realizadas em grupo.

---

<sup>11</sup>O *fact-checking* é uma checagem de fatos, isto é, um confrontamento de histórias com dados, pesquisas e registros. É uma forma de qualificar o debate público por meio da apuração jornalística. De checar qual é o grau de verdade das informações (PÚBLICA, 2017, online).

O *Data Studio* também permite o manuseio de dados em planilhas, mas seu foco é na visualização das informações, por meio de relatórios e/ou painéis informativos interativos e personalizáveis, com possibilidade de compartilhamento e colaboração. Um dos pontos mais interessantes sobre essa aplicação é a facilidade de conectar e acessar dados de diversas fontes sem a necessidade de aplicar códigos de programação.

O *Data Studio* permite rápida conexão e manipulação de conjuntos de dados provenientes de produtos do *Google Marketing Platform*, incluindo *Google Ads*, *Analytics*, *Display & Video 360* e *Search Ads 360*; produtos de consumo do Google, como o Planilhas, o *YouTube* e o *Search Console*; bancos de dados, incluindo *BigQuery*, *MySQL* e *PostgreSQL*; arquivos simples por meio do upload de arquivos *CSV* e o *Google Cloud Storage*; e também de plataformas de mídias sociais, como *Facebook*, *Reddit* e *Twitter*.

As ferramentas *Data GIF Maker*, *Tilegrams* e *Flourish* são opções para desenvolvimento de visualizações. O *Data GIF Maker* cria simples animações que podem mostrar o compartilhamento de interesse, uma comparação entre dois tópicos concorrentes. Já o *Tilegrams* é uma ótima opção para quem precisa trabalhar com visualizações de dados geográficos ou de populações por meio de mapas. Esta aplicação gera mapas feitos de blocos onde as regiões são proporcionais ao conjunto de dados em utilização, podendo representar dados demográficos com mais precisão que mapas geográficos tradicionais.

Enquanto que o *Flourish*, que é uma ferramenta nativamente jornalística, ou seja, foi desenvolvida com o objetivo de otimizar a dinâmica das redações e da produção jornalística, oferece modelos (*templates*) prontos para edição e utilização. Este último só foi disponibilizado gratuitamente graças a uma parceria entre o *Google News Lab* e os desenvolvedores da aplicação. O *Flourish* possibilita

visualizações bastante completas dos dados e com aspectos personalizáveis que não demandam o domínio de técnicas de programação/codificação.

Disponível desde 2014, o *Public Data Explorer* agrega conjuntos de dados públicos de fontes confiáveis e fornece ferramentas simples para criar visualizações sofisticadas. Por meio da plataforma é possível monitorar alterações ao longo do tempo, identificar tendências e comparar métricas. Com conjuntos de dados do Banco Mundial, Fundo Monetário Internacional, Eurostat e mais, o *Public Data Explorer* fornece estatísticas globais sobre meio ambiente, economia, trabalho, saúde, educação, pobreza e outras temáticas.

No explorador também pode-se encontrar dados mais detalhados de governos estaduais e locais, universidades e organizações sem fins lucrativos, e assim, são contempladas pesquisas mais amplas até as mais restritas com esta ferramenta. Uma das limitações do *Public Data Explorer* está relacionada à origem dos dados agregados: a maior parte das informações disponibilizadas estão em inglês, ou seja, provém de países da língua inglesa. Encontrar informações do Brasil, por exemplo, pode ser uma tarefa difícil.

Uma ferramenta recente, que não está no escopo do curso sobre Jornalismo de Dados da Google que estamos destrinchando, mas que se apresenta como um adicional ao *Google Data Explorer*, é o *Google Dataset Search*. Ainda em sua versão beta, o buscador permite a localização de conjuntos de dados armazenados em toda a Web com uma simples pesquisa por palavra-chave. A ferramenta exhibe informações sobre conjuntos de dados hospedados em milhares de repositórios na Web, de caráter público ou privado.

Na nota de apresentação da plataforma, divulgada em setembro de 2018, a pesquisadora científica da *Google Artificial Intelligence* Natasha Noy explica que, dentre os vários objetivos do lançamento do projeto, estão a criação de um ecossistema de compartilhamento de dados e a construção de um espaço onde os

cientistas possam mostrar o impacto do próprio trabalho, por meio da citação dos conjuntos de dados produzidos por eles. Noy (2018) afirma que conforme mais repositórios de conjuntos de dados adotarem o *schema.org* e padrões semelhantes para descrever tais conjuntos, a variedade e a cobertura do conteúdo encontrado no *Google Dataset Search* aumentarão.

No mundo de hoje, cientistas de várias disciplinas e um número crescente de jornalistas vivem e respiram dados. Existem muitos milhares de repositórios de dados na web, fornecendo acesso a milhões de conjuntos de dados; e governos locais e nacionais de todo o mundo publicam seus dados também. Para facilitar o acesso a esses dados, lançamos a pesquisa de conjuntos de dados, o Google Dataset Search, para que cientistas, jornalistas de dados, geeks de dados ou qualquer outra pessoa possam encontrar os dados necessários para seu trabalho e suas histórias, ou simplesmente para satisfazer sua curiosidade intelectual (NOY, 2018, on-line, tradução da autora<sup>12</sup>).

Retornando às ferramentas do curso *Data Journalism*, temos ainda a *Google Trends*. Tal ferramenta permite a visualização dos tópicos/assuntos mais recorrentes entre as pessoas comuns em tempo real que, para jornalistas, pode representar uma maneira de identificar pautas de interesse com a possibilidade de utilizar os dados da própria *Google Trends* para sustentar as reportagens.

As informações apresentadas na aplicação provêm dos volumes de pesquisas no buscador geral da Google, no Google Notícias ou no *YouTube*. A ferramenta permite a criação de filtros por regiões e países, por intervalos de tempo, além de possibilitar a comparação entre duas palavras-chave distintas, e também a geração de visualizações das informações buscadas por meio de gráficos para linhas de tempo e de mapas.

Por fim, há o *Google Surveys*, uma ferramenta de soluções em Marketing do Google com foco no desenvolvimento de pesquisas direcionadas a públicos de

---

<sup>12</sup> In today's world, scientists in many disciplines and a growing number of journalists live and breathe data. There are many thousands of data repositories on the web, providing access to millions of datasets; and local and national governments around the world publish their data as well. To enable easy access to this data, we launched Dataset Search, so that scientists, data journalists, data geeks, or anyone else can find the data required for their work and their stories, or simply to satisfy their intellectual curiosity.



interesse sobre quaisquer temáticas. Entretanto, por se tratar de um serviço voltado principalmente às empresas, o *Google Surveys* é pago.

Em contrapartida, uma alternativa gratuita, também voltada para a coleta de dados por meio de formulários, é o *Google Forms*, ou Formulários Google, que recolhe respostas e gera automaticamente gráficos baseados nas informações inseridas que podem ser consultados em tempo real. A base de dados construída por um formulário pode ser ainda analisada e manipulada pela integração com outra ferramenta já apresentada, a *Google Sheets*.

Portanto, com base no modelo sugerido no tópico anterior deste capítulo, elaboramos um guia de aplicação para as ferramentas da Google nas etapas do *workflow* do Jornalismo de Dados.

Tabela 2 - Ferramentas do Google para Etapas do Fluxo de Trabalho do JD

<b>1. OBTENÇÃO DOS DADOS</b>	<i>Public Data Explorer</i> <i>Google Trends</i> <i>Google Crisis Map</i> <i>Google Surveys</i>
<b>2. ORDENAÇÃO DOS DADOS</b>	<i>Google Sheets</i>
<b>3. ANÁLISE DOS DADOS</b>	<i>Google Sheets</i> <i>Data Studio</i>
<b>4. REPRESENTAÇÃO DOS DADOS</b>	<i>Google Sheets</i> <i>Data Studio</i> <i>Google Data GIF Maker</i> <i>Tilegrams</i> <i>Flourish</i>

Fonte: Elaborada pela autora (2019)

Por fim, a *Google News Initiative* também disponibiliza ferramentas de gestão com foco em soluções mercadológicas — a exemplo do *Google Surveys* outrora

citado — para que empresas, e/ou veículos de comunicação, trabalhem mais estrategicamente com dados. Atualmente estão disponíveis três ferramentas: *Realtime Content Insights* (RCI), *News Consumer Insights* (NCI) e *Data Maturity Benchmark* (DMB).

A RCI, que trata de informações em tempo real, possibilita que identificação *realtime* de quais artigos são mais populares entre os leitores e quais temáticas estão em alta naquela região, permitindo ainda a visualização de dados de interesse para a redação. A solução NCI atua na avaliação de desempenho da audiência (dos leitores) e sugere melhorias para ampliar os resultados.

Enquanto que a DMB gera diagnósticos sobre a posição da empresa/veículo/agência, numa escala de maturidade de dados, e também sugere maneiras de avançar nessa avaliação. Todas as ferramentas estão vinculadas ao *Google Analytics* e estão disponíveis para acesso no site da *Google News Initiative*.

### **3.3 Sobre Representação dos Dados e Infografia**

Segundo os autores que apresentamos no início deste capítulo, a última etapa de qualquer processo de produção em Jornalismo de Dados é a representação, ou visualização, dos dados (MEYER, 1991; LORENZ, 2010; BRADSHAW, 2011; ROGERS, 2013; SILVER, 2014).

Após todo o trabalho de obtenção, organização e análise dos dados, o jornalista responsável pela pauta e pela seleção do conjunto de dados utilizado na apuração precisa estudar a melhor forma de apresentar as informações e os resultados obtidos na sua observação para o seu público.

Como afirmado por Pedroza, Bezerra e Nicolau (2014, p. 1), “a premissa fundante do jornalismo é a sua capacidade de transformar informações e dados em matérias devidamente organizadas para compreensão do público” e para mais “o

jornalismo também é agente de letramento, quando guia e ensina ao leitor o caminho da apreensão de certos assuntos” (CARVALHO e SANTOS, 2014, p. 13).

Sendo assim, fundamentalmente no jornalismo dedicado aos dados, em nada adianta uma longa reportagem recheada de termos técnicos e processuais, se toda esta massa de informações não está traduzida em um formato visual, resumido, que traga significado para os dados apresentados e possibilidade de associação para o indivíduo que realiza a leitura daquele conteúdo.

Ao pesquisarmos sobre visualização de informação no universo do jornalismo, o primeiro e mais recorrente apontamento que nos retorna é sobre a produção de infografia, e quando falamos então sobre reportagens baseadas em dados é esperado o desenvolvimento de um infográfico, ou minimamente a utilização de um gráfico simples acompanhado de um texto explicativo, que traduza, e por si só, aquele conteúdo complexo que está sendo publicado.

[...] Para além de fotos, ilustrações e variações tipográficas, destaca-se como principal elemento do jornalismo visual o infográfico, formado por uma combinação de texto e imagem que se coloca como opção para permitir a compreensão de um fenômeno difícil de ser descrito por meio de uma narrativa convencional, predominantemente textual (TEIXEIRA, 2010; KANNO, 2013). Nas matérias baseadas em dados, os elementos visuais permitem construir uma visualização bem projetada, importante para evitar a desorganização de uma história complexa e auxiliar na construção de um modelo mental de um fato, tendência ou processo (MCGHEE, 2012) [...] (SALDANHA e ANDRADE, 2018, p. 9).

Há uma importante diferença entre gráficos e o infográfico ou infografia que precisamos elucidar enquanto colaboramos para o entendimento da relevância da visualização para o Jornalismo de Dados. Embora os infográficos sejam elementos comuns ao jornalismo contemporâneo e digital, estes se comportam de maneira distinta da estrutura gráfica utilizada pelo jornalismo tradicional, que costuma incluir mapas, gráficos estatísticos, diagramas e ilustrações que servem como complementos dos textos escritos.

Ricardo C. Lima (2009), em dissertação de mestrado onde trata da análise da infografia jornalística, afirma que este tipo de infografia é um meio de informação que combina os diferentes recursos utilizados pela narrativa jornalística tradicional e atuam como matérias jornalísticas independentes, não sendo necessariamente subordinados a outros textos jornalísticos. Para a pesquisadora Tattiana Teixeira — da Universidade Federal de Santa Catarina, responsável por uma obra bastante completa acerca da infografia jornalística — os infográficos reproduzem de maneira visual um acontecimento ou objeto (2010, p. 18). Entretanto, Lima reitera que:

Não devemos concluir que a infografia seja necessariamente uma simplificação de uma informação complexa, embora ela também possa ser. Sua função é contextualizar a informação para o leitor, trazendo, para primeiro plano, questões de compreensão visual e verbal muitas vezes ignoradas na tradição da linguagem verbal (LIMA, 2009, p. 32).

Os formatos de se apresentar uma infografia também estão em transformação constante, acompanhando o passo de todas as inovações tecnológicas que estão influenciando a maneira de se fazer jornalismo. A Web que anteriormente possibilitava o anexo de um simples gráfico ou infográfico estático à uma reportagem, hoje concede suporte à infografia com recursos de animação, com atualizações em tempo real, possibilitando inclusive a interatividade com o público.

## CAPÍTULO 4 – ENSINO DE JORNALISMO DE DADOS NO BRASIL

Apresentamos ao longo deste trabalho os contextos e características que envolvem o Jornalismo de Dados e seus profissionais, com argumentos baseados nas ideias dos principais pensadores desse tipo de jornalismo, como Meyer (1991), Bradshaw (2011), Bounegru (2012) e Trasel (2014).

Um dos principais pontos do debate acerca do JD — ao lado de questões como a definição concreta de um conceito, a consolidação de um fluxo e da delimitação da estrutura ferramental básica para a apuração com dados —, e sobre o qual falaremos neste capítulo final, tem a ver com as habilidades e formação necessária ao jornalista que queira direcionar sua carreira à produção de reportagens guiadas por dados.

Neste sentido, o objetivo específico desta seção é dar resposta, com base nos apontamentos teóricos que serão apresentados, a um importante questionamento: as instituições de ensino do país estão preparando os futuros jornalistas para o trabalho com dados?

*Big Data*, análise semântica, ferramentas de estruturação e análise de dados, programação: termos que, apesar de cada vez mais inseridos no vocabulário popular, ainda são explicados por conceitos densos de entendimento nem sempre imediato. Contudo, para um jornalista de dados, devem ser reconhecíveis, familiares, como quaisquer outras expressões do universo jornalístico.

Oliveira e Angeluci (2019, p. 399-340) afirmam que, devido ao contexto fundamentalmente tecnológico que vivemos hoje, exige-se dos profissionais de jornalismo, desde os mais experientes até os que ainda estão em formação, o desenvolvimento de conhecimentos específicos com foco na manipulação e tratamento de dados digitais, “em uma perspectiva de especialização jornalística fluida, multiplataforma, computacional e mais interdisciplinar”.

Os autores concluíram em 2018 uma pesquisa bastante completa acerca das habilidades e competências do jornalista de dados brasileiro, que apresentou dados baseados em entrevistas com especialistas, profissionais do mercado e professores ligados às práticas do jornalismo de dados no país. Com as informações adquiridas no estudo, Oliveira e Angeluci elaboraram uma cartografia interativa chamada *2030 Data Journalism*, estruturado com as informações da matriz de competências e habilidades também desenvolvida pelos autores.

Por fim, o estudo indicou que neste momento tão decisivo na relevância da profissão e reputação do jornalista, o jornalismo de dados vem complementar a narrativa noticiosa clássica, com mais precisão, acurácia e exatidão. Ele complementa, valida, qualifica e cria uma narrativa mais consistente, amigável, interativa e voltada ao usuário digital. É consenso entre todos os entrevistados que o jornalista com competências e habilidades em jornalismo de dados consegue aliar, de forma muito mais coerente, sua visão sociológica voltada ao interesse público a uma abordagem mais analítica e precisa, fundamentada em dados. Outro consenso é de que esse conhecimento deve ser disseminado não só entre os estudantes e jovens profissionais, como também entre professores universitários, que precisam estar capacitados de forma adequada para formar novas gerações de profissionais (OLIVEIRA e ANGELUCI, 2019, p. 414).

Destacamos também uma referência relevante sobre a formação acadêmica dos futuros profissionais de jornalismo, que é a afirmação feita por Molina e Medeiros (2017), em artigo para o Simpósio Internacional de Educação e Comunicação (SimEduC), no qual os professores trataram das transformações no perfil do jornalista em relação às observações feitas no jornal universitário digital da Universidade de Brasília (UnB), o *Campus Multimídia*.

Web semântica, HTML, CSS, linguagens de programação, algoritmos, estatísticas, big data, formatos de áudio, vídeo e imagens... Qual o limite? O jornalista do mundo de hoje lida, em maior ou menor escala e profundidade, com todos estes elementos e muito mais. Pois eles interferem diretamente no fazer jornalístico, em todas as suas etapas. Ou as escolas abraçam este novo desafio, que pode levar a uma reformulação profunda de seus currículos, ou colocará no mercado profissionais de capacidade limitada, dependentes em alto grau de competências que estarão em outras áreas (MOLINA e MEDEIROS, 2017, p. 13).

Podemos perceber que há um consenso entre as avaliações de cenários feitas pelos quatro autores citados: a formação acadêmica precisa se atualizar para acompanhar o passo das inovações constantes deste cenário contemporâneo em que estamos inseridos, e assim passar a formar para então abastecer o mercado com profissionais mais capacitados para as novas tecnologias desta Era da Informação, momento em que o dado se tornou um bem de grande valor.

Considerando todo este contexto e as referências teóricas citadas acima, assim como ao longo de todo o trabalho, apresentaremos então os resultados obtidos da busca de dados para responder ao questionamento, o objetivo específico exposto no início deste capítulo.

#### **4.1 Faculdades de Jornalismo no Brasil**

Com base no último Ranking Universitário desenvolvido pelo jornal *Folha de S. Paulo* neste ano, selecionamos as quinze melhores faculdades de Comunicação segundo a classificação nacional e analisamos os currículos de graduação em Jornalismo destas, observando a presença ou não de disciplinas obrigatórias sobre dados nas atuais matrizes curriculares disponibilizadas nos sites e portais das instituições.

O Ranking Universitário Folha (RUF) é feito anualmente pelo jornal Folha de S. Paulo desde 2012. Nele estão classificadas 197 universidades do país, públicas e privadas, considerando cinco indicadores: pesquisa, internacionalização, inovação, ensino e mercado. Além da avaliação das instituições, o RUF também apresenta uma lista dos 40 cursos de graduação com mais ingressantes no ano de cada edição.

O ranking é construído com base nas informações coletadas pela equipe da Folha em bases de dados internacionais e nacionais de trabalhos científicos, de patentes, em bases do Inep-MEC (Censo da Educação Superior e Enade), em agências estaduais e federais de fomento à ciência e em pesquisas nacionais de

opinião feitas pelo Datafolha (O QUE, 2019, online). Com base no RUF 2019, destacamos na tabela 3 o recorte dos 15 cursos de graduação em Comunicação melhores avaliados, considerando os cinco indicadores anteriormente citados.

Tabela 3 - Mapeamento com quinze maiores instituições nacionais de ensino superior  
X oferta de disciplinas de Jornalismo de Dados

RUF 2019	NOME DA INSTITUIÇÃO	TIPO	POSSUI DISCIPLINA SOBRE DADOS?	NOME DA DISCIPLINA
1º	UFRGS	Pública	Não	-
2º	UFMG	Pública	Não	-
3º	UFSC	Pública	Sim	Optativas: Jornalismo Cidadão, Jornalismo de Dados e Infografia
4º	USP	Pública	Não	-
5º	Cásper Líbero	Privada	Não	-
6º	PUCSP	Privada	Sim	Jornalismo de Dados
7º	PUCRS	Privada	Sim	Dados e Fact Checking
8º	UFRJ	Pública	Sim	Jornalismo de Dados
9º	UnB	Pública	Não	-
10º	UFPR	Pública	Sim	Jornalismo investigativo e de dados
11º	ESPM SP	Privada	Sim	Análise e Visualização de Dados; Jornalismo de Dados
12º	MACKENZIE	Privada	Sim	Jornalismo de Dados
13º	UFBA	Pública	Não	-
14º	PUC Minas	Privada	Não	-
15º	UFC	Pública	Não	-

Fonte: Elaborada pela autora (2019); com base em dados do RUF (2019)

Observamos que dentre as 15 melhores faculdades de Jornalismo do país, sete delas possuem disciplinas de Jornalismo de Dados nos seus currículos, somente uma, a Universidade Federal de Santa Catarina (UFSC), oferta três optativas sobre o tema, conforme a tabela 3. As sete restantes não possuem matérias específicas e obrigatórias sobre o assunto — considerando que nosso



escopo de busca foi direcionado a expressões e termos diretamente relacionados ao Jornalismo de Dados e que foram previamente citados neste trabalho.

Para confirmar estes resultados, foi realizada uma busca em bancos de dados sobre educação no país, sendo estes: Laboratório de Dados Educacionais (LDE), Sistema Integrado de Monitoramento do Ministério da Ciência, Tecnologia, Inovações e Comunicações (SIMMCTIC) e a *Blended Integrated Open Data*. A busca indicou que as mesmas sete faculdades do escopo anterior, pelo RUF (2019, online), são de fato, as instituições de ensino superior nacionais que oferecem alguma disciplina no âmbito do jornalismo e dados.

Para um entendimento mais dimensional destes resultados, considere que o último Censo da Educação Superior, consolidado em 2018, com dados de 2017, indicou que existem 344 instituições que oferecem o curso de graduação em Jornalismo, que contemplam todos os 26 estados, não contabilizando o Distrito Federal e cursos à distância. Do total, 58 (16,9%) são públicas e 286 (83,1%) privadas.

Ao cruzarmos estes dados com os recortes feitos anteriormente no mapeamento de oferta de disciplinas de Jornalismo de Dados, observamos que 5,1% das instituições públicas de ensino superior do Brasil oferecem disciplinas com foco em Jornalismo de Dados para 1,3% das faculdades particulares, e assim, para o total nacional de 344 faculdades, a taxa de presença de disciplinas sobre o tema é de 2,03%.

#### **4.2 Faculdades de Jornalismo no Distrito Federal**

No escopo do Distrito Federal, segundo o MEC (2018, online), existiriam 13 instituições de ensino superior com cursos de graduação em Jornalismo. Entretanto, ao realizarmos as buscas nos bancos de dados citados anteriormente, considerando que as informações do Censo — que são de 2017 — poderiam estar desatualizadas, identificamos que atualmente apenas sete instituições oferecem graduação na

modalidade presencial em Jornalismo no DF. Sobre este novo número fizemos o estudo dos currículos.

E assim, na análise feita das grades curriculares obrigatórias atuais dos cursos, identificamos que todas, exceto a Universidade de Brasília (UnB), já oferecem disciplinas obrigatórias aos graduandos na área de Jornalismo de Dados, considerando o mesmo escopo de busca explicado no tópico anterior.

Tabela 4 - Mapeamento das instituições de ensino superior com graduação em Jornalismo no DF X Oferta de disciplinas sobre Jornalismo de Dados

<b>INSTITUIÇÃO</b>	<b>POSSUI DISCIPLINA SOBRE DADOS?</b>	<b>NOME DA DISCIPLINA</b>
IESB	Sim	Jornalismo de Dados e Infografia
UDF	Sim	Infografia
UNIP	Sim	Infografia
Estácio	Sim	Jornalismo de Dados e Algoritmos
Católica	Sim	Visualização de Dados
UnB	Não	-
UnICEUB	Sim	Jornalismo de Dados e Investigativo

Fonte: Elaborado pela autora (2019)

Até o momento da conclusão desta pesquisa, apenas a UnB, que detém um dos cursos mais antigos e tradicionais de Jornalismo do país, ainda não integrou ao currículo obrigatório desta graduação, uma disciplina específica para o ensino de Jornalismo de Dados.

Entretanto, professores da Faculdade de Comunicação (FAC) da UnB têm se mobilizado para aproximar as práticas com dados — ainda que não diretamente relacionadas ao JD mas promovendo a inserção de tecnologias e inteligências

digitais — ao dia a dia dos alunos da graduação, com iniciativas ainda não permanentes (ou obrigatórias) por meio de turmas de Tópicos Especiais, a exemplo das disciplinas de Programação para Comunicadores e de Ontologia da Internet. Porém, é planejado para o primeiro semestre de 2020 a oferta de uma disciplina optativa de Jornalismo de Dados, que tratará de teoria e também de práticas. Projetos e cursos de extensão sobre o tema também estão sendo cada vez mais promovidos pelo Departamento de Jornalismo; movimentações que desenham um cenário positivo sobre os avanços que estão acontecendo e se consolidando.

A Universidade Católica de Brasília (UCB, ou somente Católica) oferece não somente uma disciplina de graduação na área de dados, mas também, desde o primeiro semestre deste ano, disponibiliza uma pós-graduação (especialização *lato sensu*) em Jornalismo de Dados, sendo a primeira iniciativa do tipo na capital do país.

#### **4.3 Apontamentos sobre os Resultados Observados**

Primeiramente, é importante destacar que as disciplinas relacionadas a dados não fazem parte das Diretrizes Curriculares estipuladas pelo Ministério da Educação para a graduação em Jornalismo. Contudo, essas diretrizes não são atualizadas desde 2013, o que pode ser considerado uma das justificativas para a falta de oferta obrigatória de disciplinas nesta área específica.

Então, com base nos dados apresentados, notamos que ainda é baixa a taxa de presença, ou a integração de disciplinas obrigatórias específicas sobre dados nos atuais currículos de graduação em Jornalismo no Brasil. Como apresentado nas tabelas 3 e 4, de maneira geral, isto demonstra que o país ainda não está formando vigorosamente profissionais preparados para lidar com esse novo tipo de apuração guiada por dados.

Essa observação pode ser relacionada com os resultados do estudo de Oliveira e Angeluci (2019) já citada neste capítulo, no qual os pesquisadores, ao

longo das entrevistas realizadas, perceberam que os profissionais atuantes hoje no mercado construíram suas competências e habilidades para lidar com dados posteriormente à graduação, com especializações e cursos, pois somente a passagem pela graduação é insuficiente para adquirir os conhecimentos específicos indispensáveis.

Outros especialistas da área, como Walter Lima Junior (2012, p.220-221), afirmam ainda que o profissional que pretende atuar no Jornalismo de Dados necessita de outras habilidades, mais voltadas à lógica e às práticas computacionais que, na visão dos autores, ainda são pouco exploradas nas universidades de Jornalismo.

Esse cenário se conecta à afirmação feita por Molina e Medeiros (2017), também já destacada anteriormente, acerca da necessidade dos centros de ensino em renovarem seus currículos para que os cursos passem a diplomar jornalistas mais completos, multidisciplinares e preparados para encarar um horizonte hiper tecnológico, no qual se comunica por meio de dados e onde se gera quantidades imensuráveis de informação em diversos formatos e todos os dias.

É importante frisar que o uso da tecnologia não vem para substituir a essência do fazer jornalístico, mas para reinventá-lo, sem abrir mão dos conceitos tradicionais e basilares que envolvem a produção de notícia, o compromisso com a verdade e a transparência, e com boas histórias adaptadas para as novas narrativas transmídia.

## CONSIDERAÇÕES

Este trabalho teve seu início com base no seguinte questionamento: por que é importante falar sobre Jornalismo de Dados? Após desenvolvermos quatro capítulos, onde procuramos apresentar assim como contribuir com os debates existentes acerca da história, da estruturação, dos apoiadores e profissionais, e da presença no assunto na ambiente acadêmico podemos então destacar alguns motivos em resposta à pergunta-problema desta monografia.

É importante falar sobre Jornalismo de Dados porque é tendência, é inovação, é emprego e é futuro. Falamos de tendência pois o tema está em constante crescimento e segue conquistando diversos espaços. Nos últimos dois anos, essa nova forma de se fazer Jornalismo ingressou de vez na rotina profissional de jornalistas em todo o mundo, desde os mais tradicionais interessados na cobertura *hardnews*, até os que já se aventuravam pelas narrativas digitais, conteúdos de entretenimento e novas mídias — como apresentado no primeiro capítulo.

Já o quesito inovação pode ser relacionado com a tecnicidade envolvida no processo, ou no fluxo de trabalho, da apuração de reportagens sobre e com dados, como citado no capítulo um e destrinchado na terceira seção do trabalho. Observamos que esse fator inovador se apresenta fortemente em duas frentes, na transformação do perfil do jornalista e nas próprias técnicas e processos específicos e fundamentais deste tipo de Jornalismo.

Quando falamos da relação do JD com emprego, consideramos também as mudanças no perfil profissional e as habilidades extra acadêmicas requeridas para a atuação na área em paralelo com as movimentações, relatadas no primeiro capítulo, que estão acontecendo ao redor do mundo por empresas/veículos de comunicação ao valorizar esse novo formato de Jornalismo, investindo e dedicando espaços com a abertura de editorias e núcleos especializados em dados dentro de suas redações.

Esse cenário significa oportunidades de trabalho para profissionais preparados, munidos das habilidades necessárias.

Por fim, falamos em futuro. Para algo que se tornou tão popular e em tão pouco tempo, seria inconsequente fazermos previsões para o que virá, para o que se tornará o Jornalismo de Dados pois este ainda está se consolidando e possivelmente, como tudo em nossa sociedade contemporânea, será desconstruído e reconstruído novamente até que, talvez, se consolide. Ainda é tudo muito novo e ainda há muito terreno a ser explorado, em questão de formatos, ferramentas e narrativas. Mas, com base em toda a pesquisa realizada para conclusão desta monografia, as próximas páginas da história do JD parecem promissoras.

Essas conexões, entre a pergunta-problema e os motivos (que também podem ser vistos como hipóteses), são indícios consideráveis do quão relevante se tornou o Jornalismo de Dados. Esse aumento de presença e influência é sustentado pelo crescimento na produção de artigos e trabalhos acadêmicos sobre o assunto, pelo surgimento de cursos, congressos e premiações para matérias jornalísticas baseadas em dados, a criação de núcleos de dados em veículos de imprensa tradicionais, assim como em veículos menores, e também a fundação de mídias independentes especializadas na apuração e desenvolvimento de conteúdos jornalísticos com dados — um sinal da aceitação do mercado.

O Jornalismo de Dados é um movimento que ganhou força e conquistou espaço e respeito de profissionais renomados assim também como de especialistas em Comunicação e Jornalismo. É disruptivo, inovador, tecnológico e possui a capacidade de lidar e trabalhar com problemas característicos desta Era da Informação, como a produção constante e exponencial de dados pela humanidade (*Big Data*), onde há espaço para atuação de jornalistas de dados.

Destacamos ainda que o JD pode ser uma resposta, uma solução para a retomada da credibilidade da mídia perante a sociedade, para o restabelecimento da

importância do bom e credível jornalismo que se preocupa em trazer informação de qualidade para o seu público. E que pode ainda desmistificar a afirmação, que tem ser tornado cada vez mais popular, sobre o jornalismo estar morrendo. O jornalismo está se transformando, se transmutando para abarcar as mudanças das plataformas, das mídias, dos formatos e dos modelos de trabalho.

Neste ponto, surgem também questionamentos sobre as habilidades e competências necessárias para atuar como jornalista numa sociedade altamente tecnológica, que gera quantidades imensuráveis de dados digitais, num panorama de produção exacerbada de informação, em tempo real, em diversos meios e em uma velocidade incomparável, propiciada pelo advento da banda larga.

Estaríamos nós, jornalistas, preparados para entender e lidar com essa nova perspectiva? Os currículos disciplinares dos cursos superiores de Jornalismo estão suficientemente atualizados para preparar as próximas gerações de jornalistas para esse cenário? Segundo os resultados de nossa análise, ainda não. Entretanto, o que observamos ainda é positivo pois, mesmo que em passos curtos, devemos reconhecer que temos avançado no objetivo de compreender todas estas transformações que têm afetado a área de atuação profissional.

E como parte desse processo de compreensão das mudanças, talvez seja necessária uma reestruturação das matrizes curriculares. Os currículos de graduação em Jornalismo existentes hoje são ainda majoritariamente direcionados à formação tradicional, com foco em disciplinas obrigatórias e conteúdos também tradicionais, que não dialogam suficientemente com a nova realidade do mercado de trabalho.

Concluimos estas considerações com uma expectativa — com base no ritmo de expansão, de adaptação e aderência que observamos nas pesquisas: o Jornalismo de Dados continuará crescendo. Acompanhando o passo do surgimento de novas tecnologias e ferramentas, pode-se esperar a curto prazo uma forte

integração do uso de inteligência artificial na coleta de informações, e também na produção de notícias, assim como o surgimento de novos modelos de negócio de caráter colaborativo e multidisciplinar para empresas de comunicação, que podem transformar ainda mais os fluxos de trabalho e o perfil do jornalista.



## REFERÊNCIAS

ABRAHAM, Rafael. O Novo Novo Jornalismo. **O Estado de S. Paulo**, São Paulo, 29 de set. de 2011. Blog Em Foca. Disponível: <<https://brasil.estadao.com.br/blogs/em-foca/o-novo-novo-jornalismo/>>. Acesso em: 02 de nov. de 2019.

ALBUQUERQUE, Afonso de. As três faces do quarto poder. In: MIGUEL, Luís Felipe; BIROLI, Flávia. **Mídia, representação e democracia**. São Paulo: Hucitec, 2010. 301 p. ISBN 9788579700316.

ALBRECHT, Rogéria Fernandes; OHIRA, Maria Lourdes Blatt. **Bases de dados: metodologia para seleção e coleta de documentos**. Revista ACB, Santa Catarina, v. 5, n. 5, p. 131-144, ago. 2000. ISSN 1414-0594. Disponível em: <<https://revistaacb.emnuvens.com.br/racb/article/view/347/411>>. Acesso em: 02 nov. 2019.

ALMEIDA, Maurício Barcellos. **Uma introdução ao XML, sua utilização na Internet e alguns conceitos complementares**. Ci. Inf., Brasília, v. 31, n. 2, p. 5-13, maio/ago. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12903>> Acesso em: 2 nov. 2019.

A WHOLE PRESIDENTIAL CAMPAIGN CATEGORIZED. **Data Journalism Awards**, 2019. Disponível em: <<https://datajournalismawards.org/projects/a-whole-presidential-campaign-categorized/>> Acesso em: 31 out. 2019

BARBOSA, Suzana (org.). **Jornalismo Digital de Terceira Geração**. Covilhã: Labcom – Universidade da Beira Interior, 2007a. 180 p. ISBN 978-972-8790-73-8. Disponível em: <[http://www.labcomifp.ubi.pt/ficheiros/barbosa\\_suzana\\_jornalismo\\_digital\\_terceira\\_geracao.pdf](http://www.labcomifp.ubi.pt/ficheiros/barbosa_suzana_jornalismo_digital_terceira_geracao.pdf)>. Acesso em: 25 out. 2019.

BARBOSA, Suzana. **Jornalismo digital e bases de dados: mapeando conceitos e funcionalidades**. LIVRO DE ACTAS – 4º Congresso da SOPCOM, Portugal, p. 1310-1321, 2007b. Disponível em: <<http://www.bocc.ubi.pt/pag/barbosa-suzana-jornalismo-digital-bases-dados.pdf>>. Acesso em: 3 nov. 2019.

BARBOSA, Suzana. **O que é Jornalismo Digital em Base de Dados**. 15º Encontro Anual da COMPÓS: Associação Nacional dos Programas de Pós-Graduação em Comunicação,

UNESP, Bauru, p. 1-15, 2006. Disponível em: <[https://www.academia.edu/15328433/O\\_QUE\\_%C3%89\\_JORNALISMO\\_DIGITAL\\_EM\\_BASES\\_DE\\_DADOS\\_1](https://www.academia.edu/15328433/O_QUE_%C3%89_JORNALISMO_DIGITAL_EM_BASES_DE_DADOS_1)>. Acesso em: 25 out. 2019.

BARR, Caelainn; CHALABI, Mona; EVERSHED, Nick. A decade of the Datablog: 'There's a human story behind every data point'. **The Guardian**. [S. l.], 23 de mar. de 2019. Disponível em:

<<https://www.theguardian.com/membership/datablog/2019/mar/23/a-decade-of-the-datablog-theres-a-human-story-behind-every-data-point>> Acesso em: 31 out. 2019.

BRADSHAW, Paul. The inverted pyramid of data journalism. **Online Journalism Blog**, 2011. Disponível em: <<https://onlinejournalismblog.com/2011/07/07/the-inverted-pyramid-of-data-journalism/>>. Acesso em: 3 nov. 2019.

BELOCHIO, Vivian de Carvalho. **Jornalismo Colaborativo em Redes Digitais: Estratégia Comunicacional no Ciberespaço - O caso de ZeroHora.com**. Orientador: Dra. Luciana Mielniczuk. 2009. 215 p. Dissertação de Mestrado (Mestre em Comunicação) - Universidade Federal de Santa Maria, Santa Maria, RS, 2009. Disponível em: [https://www.researchgate.net/profile/Vivian\\_Belochio2/publication/284183944\\_Jornalismo\\_colaborativo\\_em\\_redes\\_digitais\\_estrategia\\_comunicacional\\_no\\_ciberespaco\\_O\\_caso\\_de\\_Zero\\_Horacom/links/564f0f3f08ae4988a7a7f2eb/Jornalismo-colaborativo-em-redes-digitais-estrategia-comunicacional-no-ciberespaco-O-caso-de-Zero-Horacom.pdf](https://www.researchgate.net/profile/Vivian_Belochio2/publication/284183944_Jornalismo_colaborativo_em_redes_digitais_estrategia_comunicacional_no_ciberespaco_O_caso_de_Zero_Horacom/links/564f0f3f08ae4988a7a7f2eb/Jornalismo-colaborativo-em-redes-digitais-estrategia-comunicacional-no-ciberespaco-O-caso-de-Zero-Horacom.pdf). Acesso em: 2 nov. 2019.

CANAVILHAS, João. **Webjornalismo: Da pirâmide invertida à pirâmide deitada**. Universidade da Beira Interior, Portugal, 2006. Disponível em: <<http://www.bocc.ubi.pt/pag/canavilhas-joao-webjornalismo-piramide-invertida.pdf>>. Acesso em: 15 dez. 2019.

CARVALHO, Alessandra Pinto de; SANTOS, Andreza Almeida. **Entremeios do processo de leitura de infográficos e visualização de dados no jornalismo impresso**. Intercom, Paraná, 2014. Disponível em: <<http://www.intercom.org.br/sis/2014/resumos/R9-1560-1.pdf>>. Acesso em: 3 nov. 2019.

CASANUEVA, Juan Manuel. **Jornalismo e agenda de dados abertos da América Latina**. IJNet - Rede de Jornalistas Internacionais, 2015. Disponível em:

<<https://ijnet.org/pt-br/story/jornalismo-e-agenda-de-dados-abertos-da-am%C3%A9rica-latina>>. Acesso em: 3 nov. 2019.

CASTELLS, Manuel. A rede é a mensagem. In: **A Galáxia da Internet: Reflexões sobre a Internet, os negócios e a sociedade**. Rio de Janeiro: Zahar, 2003. p. 7-8. ISBN 978-85-7110-740-3. Disponível em: <[https://zahar.com.br/sites/default/files/arquivos/trecho\\_-\\_a\\_galaxia\\_da\\_internet.pdf](https://zahar.com.br/sites/default/files/arquivos/trecho_-_a_galaxia_da_internet.pdf)> Acesso em: 25 out. 2019.

CETIC. **PRINCIPAIS RESULTADOS - TIC DOMICÍLIOS 2018**. 28 ago. 2019. Apresentação em Slides. Disponível em: <[https://cetic.br/media/analises/tic\\_domicilios\\_2018\\_coletiva\\_de\\_imprensa.pdf](https://cetic.br/media/analises/tic_domicilios_2018_coletiva_de_imprensa.pdf)>. Acesso em: 3 nov. 2019.

CIANCONI, Regina de Barros. **Banco de dados de acesso público**. Ciência da Informação, Brasília, v. 16, n. 1, June 1987. Disponível em: <<http://revista.ibict.br/ciinf/article/view/271/271>>. Acesso em: 03 nov. 2019.

CORREIA, Marcos Balster Fiore. Dados, Informação e Conhecimento. In: CORREIA, Marcos Balster Fiore. **A Comunicação de Dados Estatísticos por intermédio de Infográficos: Uma Abordagem Ergonômica**. Orientador: Anamaria de Moraes. 2009. Dissertação de Mestrado (Mestre em Artes) - PUC-Rio, Rio de Janeiro, 2009. p. 463. Disponível em: <[https://www.maxwell.vrac.puc-rio.br/14038/14038\\_4.PDF](https://www.maxwell.vrac.puc-rio.br/14038/14038_4.PDF)>. Acesso em: 2 nov. 2019.

COX, M. **The Development of Computer-Assisted Reporting**. Association for Education in Journalism and Mass Communication, Southeast Colloquium, University of North Carolina, Chapel Hill, 2000. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.631.6220&rep=rep1&type=pdf>>. Acesso em: 25 out. 2019.

DATASUS. **GLOSSÁRIO: SISTEMA GERENCIADOR DE BANCO DE DADOS (SGBD)**, c2019. MAD – Metodologia de Administração de Dados. Disponível em: <<http://datasus.saude.gov.br/apoio-mad/glossario-mad-menu/963-sistema-gerenciador-de-banco-de-dados-sgbd>>. Acesso em: 3 nov. 2019.

RANKING de Cursos de Graduação - Comunicação. Ranking Universitário Folha 2019. **Folha de S. Paulo**, São Paulo, 2019. Disponível: <<https://ruf.folha.uol.com.br/2019/ranking-de-cursos/comunicacao/>>. Acesso em: 15 dez. 2019.

FOLHA LANÇA SITE DE DADOS ABERTOS FOCADO NAS ELEIÇÕES DE SÃO PAULO. **Folha de S. Paulo**. São Paulo, 20 de ago. de 2012. Poder. Disponível em: <<https://www1.folha.uol.com.br/poder/2012/08/1139892-folha-lanca-site-de-dados-abertos-focado-nas-eleicoes-de-sao-paulo.shtml>> Acesso em: 31 out. 2019

GARRISON, Bruce. **Computer-Assisted Reporting**. 2. ed. Mahwah: Lawrence Erlbaum Associates, 1998. Disponível em: <<https://www.questia.com/read/13647227/computer-assisted-reporting>>. Acesso em: 25 out. 2019.

GARRISON, Bruce. **Diffusion of online information technologies in newspaper newsrooms**. JOURNALISM AND NEW TECHNOLOGIES, University of Miami, p. 221–239, Agosto 2001. DOI 10.1177/146488490100200206. Disponível em: <[https://www.researchgate.net/publication/249689747\\_Diffusion\\_of\\_online\\_information\\_technologies\\_in\\_newspaper\\_newsrooms](https://www.researchgate.net/publication/249689747_Diffusion_of_online_information_technologies_in_newspaper_newsrooms)>. Acesso em: 25 out. 2019.

GANTZ, John; REINSEL, David. **THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows and Biggest Growth in the Far East**. IDC iView, Massachusetts, p. 1-16, 2012. Disponível em: <<https://pt.slideshare.net/arms8586/the-digital-universe-in-2020>>. Acesso em: 3 nov. 2019.

GERHARDT, Tatiana Engel; SOUZA, Aline Corrêa de (org.). **Métodos de Pesquisa**. Porto Alegre: Editora da UFRGS, 2009. 120 p. ISBN 978-85-386-0071-8. Disponível em: <<http://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf>>. Acesso em: 10 nov. 2019.

GRANDIN, Felipe Rodrigues. **Jornalismo guiado por dados como forma contemporânea de produção de sentido**. Intercom, Paraná, 2014a. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2014/resumos/R9-2519-1.pdf>>. Acesso em: 3 nov. 2019.

GRANDIN, Felipe Rodrigues. **A contribuição do jornalismo guiado por dados para a criação de valor nas organizações jornalísticas**. Orientador: Profª. Draª. Sonia Virginia Moreira. 2014b. 123 p. Dissertação de Mestrado (Mestre em Comunicação) - Universidade

do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2014. Disponível em: <http://www.ppgcom.uerj.br/wp-content/uploads/Disserta%C3%A7%C3%A3o-Felipe-Rodrigues-Grandin.pdf>>. Acesso em: 3 nov. 2019.

GRAY, Jonathan; BOUNEGRU, Liliana; CHAMBERS, Lucy (ed.). **The Data Journalism Handbook**. Sebastopol: O'Reilly Media, 2012. 238 p. ISBN 9781449330064. Disponível em: <https://datajournalism.com/read/handbook/one>>. Acesso em: 25 out. 2019.

HALPER, Fern; KRISHNAN, Krish. **TDWI Big Data Maturity Model Guide: Interpreting Your Assessment Score**. Washington: TDWI (The Data Warehousing Institute), 2013. 20 p. Disponível em: <https://tdwi.org/whitepapers/2013/10/tdwi-big-data-maturity-model-guide.asp>>. Acesso em: 3 nov. 2019.

HISTÓRIA DA FOLHA. **Folha de S. Paulo**. São Paulo, [s.d.]. Disponível em: [https://www1.folha.uol.com.br/institucional/historia\\_da\\_folha.shtml?fill=4](https://www1.folha.uol.com.br/institucional/historia_da_folha.shtml?fill=4)> Acesso em: 31 out. 2019

HOLOVATY, Adrian. **A fundamental way newspaper sites need to change**. [S. l.: s. n.], 6 set. 2006. Disponível em: <http://www.holovaty.com/writing/fundamental-change/>>. Acesso em: 25 out. 2019.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Sinopse Estatística da Educação Superior 2018**. Brasília: Ministério da Educação (MEC), Inep, 2019. Disponível em: <http://portal.inep.gov.br/web/guest/sinopses-estatisticas-da-educacao-superior>>. Acesso em 6 nov. 2019.

ISOTANI, Seiji; BITTENCOURT, Ig lbert. **Dados abertos conectados**. São Paulo: Novatec, 2015. 177 p. ISBN 9788575224496. Disponível em: <https://ceweb.br/publicacao/livro-dados-abertos/>>. Acesso em: 2 nov. 2019.

KEMP, Simon. **DIGITAL IN 2018: WORLD'S INTERNET USERS PASS THE 4 BILLION MARK**. We Are Social, 2018. Disponível em: <https://wearesocial.com/blog/2018/01/global-digital-report-2018>> Acesso em: 3 nov. 2019.

KNUTH, Donald Ervin. Algorithms, Programs and Computer Science. *In: Selected Papers on Computer Science*. Stanford: CSLI Publications, 1996. p. 1-4. ISBN 1881526917.

LANEY, Doug. **3D Data Management: Controlling Data Volume, Velocity and Variety**. Gartner Blog Network, Stamford, p. 1-4, 6 fev. 2001. Disponível em: <<https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>> Acesso em: 2 nov. 2019.

LIMA, Patrícia. **Jornalismo em tempos de dados digitais: Reconfigurações das práticas e da identidade jornalística**. Intercom, Curitiba, 2017. Disponível em: <<http://portalintercom.org.br/anais/nacional2017/resumos/R12-0592-2.pdf>>. Acesso em: 3 nov. 2019.

LIMA, Ricardo Oliveira da Cunha. **Análise da Infografia Jornalística**. Orientador: Prof. Doutor André Soares Monat. 2009. 145 p. Dissertação de mestrado (Mestre em Design) - Escola Superior de Desenho Industrial, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2009. Disponível em: <<https://www.researchgate.net/publication/231220329>>. Acesso em: 3 nov. 2019.

LIMA, Telma Cristiane Sasso de; MIOTO, Regina Célia Tamasso. **Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica**. Rev. Katál, Florianópolis, v. 10, p. 37-45, 2007. Disponível em: <<http://www.scielo.br/pdf/rk/v10nspe/a0410spe.pdf>>. Acesso em: 10 nov. 2019.

LIMA JUNIOR, Walter Teixeira. **Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados**. Estudos em Comunicação, Universidade Metodista de São Paulo, n. 12, p. 207-222, Dezembro 2012. Disponível em: <[https://www.researchgate.net/publication/233954587\\_Big\\_Data\\_Jornalismo\\_Computacional\\_e\\_Data\\_Journalismestrutura\\_pensamento\\_e\\_pratica\\_profissional\\_na\\_Web\\_de\\_dados](https://www.researchgate.net/publication/233954587_Big_Data_Jornalismo_Computacional_e_Data_Journalismestrutura_pensamento_e_pratica_profissional_na_Web_de_dados)>. Acesso em: 25 out. 2019.

LORENZ, Mirko. **Data-Driven Journalism**. Abril, 2010. Apresentação em Slides. Disponível em: <<https://pt.slideshare.net/mirkolorenz/datadriven-journalism-presentation-gijc-geneva-april-2010>>. Acesso em: 3 nov. 2019.

MARTINS, César Silva; SIMÕES, Paulo; SÁ, Jorge Vaz de Oliveira e. **Uma arquitetura moderna de dados: Um caso de teste**. Centro ALGORITMI, Escola de Engenharia. Universidade do Minho, Portugal, 2014. Disponível em:

<[https://repositorium.sdum.uminho.pt/bitstream/1822/31369/1/paper\\_cpasi\\_2014\\_arquitetura\\_moderna\\_vf.pdf](https://repositorium.sdum.uminho.pt/bitstream/1822/31369/1/paper_cpasi_2014_arquitetura_moderna_vf.pdf)>. Acesso em: 2 nov. 2019.

MARTINHO, Ana Pinto. **Jornalismo de Dados: caracterização e fluxos de trabalho**. Exedra Journal: Comunicação e Ciências Empresariais, Lisboa, n. 9, p. 64-73, 2014. Disponível em: <<http://docplayer.com.br/871998-Jornalismo-de-dados-caraterizacao-e-fluxos-de-trabalho.html>>. Acesso em: 25 out. 2019.

MAZZOTE, Natália. **As novas fronteiras do jornalismo de dados no Brasil**. Medium, 2016. O jornalismo no Brasil em 2017. Disponível em: <<https://jornalismonobrasil2017.com/as-novas-fronteiras-do-jornalismo-de-dados-no-brasil-cddeede4fb8d>>. Acesso em: 3 nov. 2019.

MIELNICZUK, Luciana; TRÄSEL, Marcelo. **JORNALISMO GUIADO POR DADOS COMO INOVAÇÃO PROFISSIONAL E SEUS DESAFIOS PARA A EDUCAÇÃO**. Contemporanea - Revista de Comunicação e Cultura, Bahia, v. 15, ed. 2, p. 609-629, 2017. DOI <http://dx.doi.org/10.9771/1809-9386contemporanea.v15i2.21510>. Disponível em: <<https://portalseer.ufba.br/index.php/contemporaneaposcom/article/view/21510>>. Acesso em: 7 nov. 2019.

MOLINA, Fernando; MEDEIROS, Benedito. **O PERFIL DO JORNALISTA 3.0: Novas competências necessárias o jornalismo no século XXI**. Anais do 8º SimEduC, Aracaju, p. 1-15, 2017. Disponível em: <[https://www.researchgate.net/publication/322131527\\_O\\_PERFIL\\_DO\\_JORNALISTA\\_30\\_Novas\\_competencias\\_necessarias\\_ao\\_jornalismo\\_no\\_seculo\\_XXI](https://www.researchgate.net/publication/322131527_O_PERFIL_DO_JORNALISTA_30_Novas_competencias_necessarias_ao_jornalismo_no_seculo_XXI)>. Acesso em: 7 nov. 2019.

MOURA, Catarina. **Jornalismo na era Slashdot**. Biblioteca Online de Ciências da Comunicação, 2002. Disponível em: <<http://bocc.ubi.pt/pag/moura-catarina-jornalismo-slashdot.html>>. Acesso em: 2 nov. 2019.

NOGUEIRA, Luis Carlos. **Slashdot, comunidade de palavra**. Biblioteca Online de Ciências da Comunicação, 2002. Disponível em: <[http://www.bocc.ubi.pt/pag/nogueira-luis-slashdot\\_texto.pdf](http://www.bocc.ubi.pt/pag/nogueira-luis-slashdot_texto.pdf)>. Acesso em: 2 nov. 2019.

NOY, Natasha. Making it easier to discover datasets. **Google Blog**, 2018. The Keyword. Disponível em: <<https://www.blog.google/products/search/making-it-easier-discover-datasets/>>. Acesso em: 3 nov. 2019.

O QUE é o Ranking Universitário Folha. **Folha de S.Paulo**, São Paulo, 07 de out. de 2019. Disponível em: <<https://rnf.folha.uol.com.br/2019/noticias/o-que-e-o-ranking-universitario-folha.shtml>>. Acesso em 8 dez. de 2019.

OLIVEIRA, Ana Paula Borges de; ANGELUCI, Alan César Belo. **COMPETÊNCIAS E HABILIDADES NO JORNALISMO DE DADOS: Percepções sobre o perfil do profissional brasileiro**. SBPJor, Brasília, v. 15, ed. 1, p. 398-417, 2019. DOI 10.25200/BJR.v15n2.2019.1141. Disponível em: <<https://drive.google.com/file/d/0B0oyCCDSY5WLLTdachHprdlRTZHpKV2dXVTNyb3lPeXFWZnJn/view?usp=sharing>>. Acesso em: 7 nov. 2019.

PALLARO, Bianca. Open Data Journalism for Change. **La Nación**. Argentina, 4 de abr. de 2019. Disponível em: <<http://blogs.lanacion.com.ar/projects/data/open-data-journalism-for-change-3/>> Acesso em: 31 out. 2019.

PEDROZA, Natan; BEZERRA, Ed Porto; NICOLAU, Marcos. **Os Recursos da Infografia na Prática do Webjornalismo**. Intercom, Manaus, 2013. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2013/resumos/R8-1586-1.pdf>>. Acesso em: 3 nov. 2019.

PIRES, Marco Túlio. Evoluir, elevar, empoderar: como o Google News Initiative ajuda o jornalismo a prosperar na era digital. **Think With Google**. [S. l.], jul. de 2019. Disponível em: <<https://www.thinkwithgoogle.com/intl/pt-br/advertising-channels/novas-tecnologias/evoluir-elevar-empoderar-como-o-google-news-initiative-ajuda-o-jornalismo-prosperar-na-era-digital/>>. Acesso em: 31 out. 2019.

POSSAMAI, Ana Júlia; SOUZA, Vitória Gonzatti de. **TRANSPARÊNCIA E DADOS ABERTOS NO GOVERNO FEDERAL: POSSIBILIDADES E DESAFIOS A PARTIR DA LEI DE ACESSO À INFORMAÇÃO**. Anais do Encontro Nacional de Ensino e Pesquisa do Campo de Públicas (ENEPCP), Rio Grande do Sul, v. 2, n. 2, p. 688-715, 2017. Disponível em:



<[https://www.anepcp.org.br/acp/anaisenepcp/20180723152454\\_35\\_Transparencia\\_e\\_dados\\_abertos\\_Ana\\_Possamai.pdf](https://www.anepcp.org.br/acp/anaisenepcp/20180723152454_35_Transparencia_e_dados_abertos_Ana_Possamai.pdf)>. Acesso em: 3 nov. 2019.

RAMAKRISHNAN, Raghu; GEHRKE, Johannes. **Sistemas de Gerenciamento de Banco de Dados**. Tradução: Célia Taniwake. 3. ed. Brasil: Mc Graw Hill, 2008. 905 p. ISBN 9788577260270.

ROCHA, Joana D'arque da Costa. **Análise de Uma Ferramenta Utilizada Como Repositório De Dados Abertos Governamentais Para Objetos De Aprendizagem Fragmentados**. Orientador: Profª. Esp. Rossana de Paula Junqueira Almeida. 2014. 45 p. Monografia de Graduação (Licenciatura em Computação) - Instituto Federal de Educação, Ciência e Tecnologia do Sertão Pernambucano, Petrolina, 2014. Disponível em: <[encurtador.com.br/dP179](http://encurtador.com.br/dP179)>. Acesso em: 3 nov. 2019.

ROGERS, Simon. A data journalism workflow. **Simon Rogers - Data Journalism and Other Curiosities**, 2013. Disponível em: <<https://simonrogers.net/2013/01/27/a-data-journalism-workflow/>>. Acesso em: 3 nov. 2019.

ROGERS, Simon. The five Ws of data journalism. **Simon Rogers - Data Journalism and Other Curiosities**, 2014. Disponível em: <<https://simonrogers.net/2014/10/16/the-five-ws-of-data-journalism/>>. Acesso em: 3 nov. 2019.

ROGERS, Simon; SCHWABISH, Jonathan; BOWERS, Danielle. **Data Journalism in 2017: The current state and challenges facing the field today**. Google News Lab, [S. l.], p. 1-24, Set. 2017. Disponível em: <https://newslab.withgoogle.com/assets/docs/data-journalism-in-2017.pdf>. Acesso em: 25 out. 2019.

SALDANHA, Felipe Gustavo Guimarães; ANDRADE, Ivanise Hilbig de. **Visualização de Dados no Ensino de Jornalismo em Defesa dos Direitos Humanos: Produção de Infográficos no Blog “Minorias de Aço”**. V Seminário de Pesquisa em Jornalismo Investigativo da Abraji, São Paulo, 2018. Disponível em: <[https://www.abraji.org.br/seminario/PDF/5/FELIPESALDANHA\\_IVANISEA\\_NDRADE\\_Abraji\\_2018.pdf](https://www.abraji.org.br/seminario/PDF/5/FELIPESALDANHA_IVANISEA_NDRADE_Abraji_2018.pdf)>. Acesso em: 3 nov. 2019.

SANTO, Ana Espírito. **Categorização e Análise de Dados Não Estruturados: O Caso dos Debates Parlamentares**. Orientador: Dr. Miguel Neto. 2009. 117 p. Dissertação de Mestrado (Mestre em Estatística e Gestão de Informação) - Universidade Nova de Lisboa, Lisboa, 2009. Disponível em: <<https://run.unl.pt/bitstream/10362/2674/1/TEGI0247.pdf>>. Acesso em: 2 nov. 2019.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S.; VIEIRA, Daniel. **Sistema de Banco de Dados**. 6. ed. Brasil: Elsevier, 2012. 904 p. ISBN 9788535245356.

SILVER, Nate. What the Fox Knows. **FiveThirtyEight**, 2014. Disponível em: <<https://fivethirtyeight.com/features/what-the-fox-knows/>>. Acesso em: 3 nov. 2019.

SOBRE. **O Estado de S. Paulo**. São Paulo, c2014. Disponível em: <<http://blog.estadaodados.com/sobre/>> Acesso em: 31 out. 2019.

SOBRE O NEXO. **Nexo Jornal**, c2019. Disponível em: <<https://www.nexojornal.com.br/about/Sobre-o-Nexo>> Acesso em: 31 out. 2019

SOBRE O VOLT. **Volt Data Lab**, [s.d.]. Disponível em: < <https://www.voltdata.info/sobre>> Acesso em: 31 out. 2019

TAKAI , Osvaldo Kotaro; ITALIANO , Isabel Cristina; FERREIRA, João Eduardo. **Introdução a Banco de Dados**. 2005. 124 p. Apostila para Disciplinas de introdução a Sistemas Banco de Dados - USP, São Paulo, 2005. Disponível em: <<https://docplayer.com.br/1373957-Introducao-a-banco-de-dados-o-k-takai-i-c-italiano-j-e-ferr-eira-1-introducao-a-banco-de-dados.html>>. Acesso em: 2 nov. 2019.

TAFFNER, Ricardo; LIMA, Rafaela; MARCHESINI, Lucas. Metrôpoles investe em jornalismo de dados e lança nova editoria. **Metrôpoles**, 2019. (M)Dados. Disponível em: <<https://www.metropoles.com/brasil/imprensa/metropoles-investe-em-jornalismo-de-dados-e-lanca-nova-editoria>> Acesso em: 31 out. 2019

TARGINO, Maria das Graças. **Jornalismo Cidadão: Informa ou deforma?**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), 2009. 260 p. ISBN 978-85-7013-065-5.

TEIXEIRA, Tattiana. **Infografia e Jornalismo: Conceitos, Análises e Perspectivas**. Salvador: EDUFBA, 2010. 126 p. ISBN 978-85-232-0746-5.

THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. **Creativante**, 2013. Disponível em: <<http://creativante.com/new/index.php/2013-02-03-19-36-05/2013-02-04-18-19-49/149-the-digital-universe-in-2020-big-data-bigger-digital-shadows-and-biggest-growth-in-the-far-east>> Acesso em: 3 nov. 2019.

TRÄSEL, Marcelo. **Entrevistando Planilhas: Estudo das crenças e do ethos de um grupo de profissionais de Jornalismo Guiado por Dados no Brasil**. Orientador: Dr. Francisco Rüdiger. 2014. 315 p. Tese (Doutorado em Comunicação Social) - Programa de Pós-Graduação em Comunicação Social, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2014. Disponível em: <<http://hdl.handle.net/10923/6841>>. Acesso em: 25 out. 2019.

TRÄSEL, Marcelo. **Jornalismo Guiado por Dados: características definidoras e uma proposta de formulação do conceito**. SBPJor - 15º Encontro Nacional de Pesquisadores em Jornalismo, São Paulo, 2017. Disponível em: <<http://sbpjor.org.br/congresso/index.php/sbpjor/sbpjor2017/paper/viewFile/794/464>>. Acesso em 4 nov. 2019.

VENTURA, Mariane Pires. **Jornalismo de Dados como diferencial: o caso do Nexô**. Revista Pauta Geral: Estudos em Jornalismo, Joinville, p. 240-245, 2018. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/6773400.pdf>. Acesso em: 25 out. 2019.

XAVIER, Rodolfo Coutinho Moreira; COSTA, Rubenildo Oliveira da. **Relações mútuas entre informação e conhecimento: o mesmo conceito?**. Ci. Inf., Brasília, v. 39, n. 2, p. 75-83, Aug. 2010. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652010000200006&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652010000200006&lng=en&nrm=iso)>. Acesso em: 2 nov. 2019.

ZANOTTI, Carlos Alberto. **Jornalismo colaborativo, gêneros jornalísticos e critérios de noticiabilidade**. Revista Comunicação Midiática, Campinas, v. 5, ed. 1, p. 28 - 41, 2010. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=3989379>. Acesso em: 2 nov. 2019.

WATSON MARKETING. **10 Key Marketing Trends for 2017 and Ideas for Exceeding Customer Expectations**. IBM MARKETING CLOUD, 2017. Apresentação em Slides. Disponível em: <<https://business-iq.net/whitepapers/1766-10-key-marketing-trends-for-2017-and-ideas-for-exceeding-customer-expectations?v=cmoatwork>>. Acesso em: 3 nov. 2019.

WHAT IS OPEN DATA?. **Open Data Handbook**, c2019. Guide. Disponível em: <<http://opendatahandbook.org/guide/en/what-is-open-data/>>. Acesso em: 3 nov. 2019.